

**MINISTÉRIO DA DEFESA
EXÉRCITO BRASILEIRO
DEPARTAMENTO DE CIÊNCIA E TECNOLOGIA
INSTITUTO MILITAR DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE DEFESA**

RIGEL PROCÓPIO FERNANDES

**ACOUSTIC-BASED DRONE LOCALIZATION AND DOA
ESTIMATION FOR HIGHLY NOISY ENVIRONMENTS**

**RIO DE JANEIRO
2025**

RIGEL PROCÓPIO FERNANDES

ACOUSTIC-BASED DRONE LOCALIZATION AND DOA
ESTIMATION FOR HIGHLY NOISY ENVIRONMENTS

Tese apresentada ao Programa de Pós-graduação em Engenharia de Defesa do Instituto Militar de Engenharia, como requisito parcial para a obtenção do título de Doutor em Engenharia de Defesa.

Orientador(es): José Antonio Apolinário Jr., D.Sc.
Julio Cesar Duarte, D.Sc.
José Manoel de Seixas, D.Sc.

Rio de Janeiro
2025

©2025

INSTITUTO MILITAR DE ENGENHARIA

Praça General Tibúrcio, 80 – Praia Vermelha

Rio de Janeiro – RJ CEP: 22290-270

Este exemplar é de propriedade do Instituto Militar de Engenharia, que poderá incluí-lo em base de dados, armazenar em computador, microfilmар ou adotar qualquer forma de arquivamento.

É permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do(s) autor(es) e do(s) orientador(es).

Fernandes, Rigel Procópio.

Acoustic-based drone localization and DOA estimation for highly noisy environments / Rigel Procópio Fernandes. – Rio de Janeiro, 2025.

109 f.

Orientador(es): José Antonio Apolinário Jr., Julio Cesar Duarte e José Manoel de Seixas.

Tese (doutorado) – Instituto Militar de Engenharia, Engenharia de Defesa, 2025.

1. VANT; 2. Drone; 3. DOA; 4. Sistemas inteligentes. i. Antonio Apolinário Jr., José (orient.) ii. Cesar Duarte, Julio (orient.) iii. Manoel de Seixas, José (orient.) iv. Título

RIGEL PROCÓPIO FERNANDES

**Acoustic-based drone localization and DOA
estimation for highly noisy environments**

Tese apresentada ao Programa de Pós-graduação em Engenharia de Defesa do Instituto Militar de Engenharia, como requisito parcial para a obtenção do título de Doutor em Engenharia de Defesa.

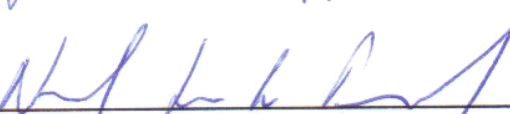
Orientador(es): José Antonio Apolinário Jr., Julio Cesar Duarte e José Manoel de Seixas.

Aprovada em 4 de julho de 2025, pela seguinte banca examinadora:

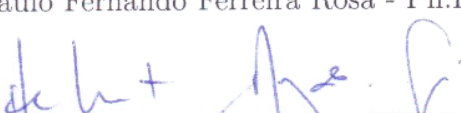

Prof. José Antonio Apolinário Jr. - D.Sc. do IME - Presidente


Prof. Julio Cesar Duarte - D.Sc. do IME


Prof. João Abdalla Ney da Silva - D.Sc. do CTEEx


Prof. Natanael Nunes de Moura Junior - D.Sc. da COPPE/UFRJ


Prof. Paulo Fernando Ferreira Rosa - Ph.D. do IME


Prof. Hebert Azevedo Sá - Ph.D. do IME

Rio de Janeiro

2025

Dedico este trabalho à minha amável esposa Paola!

ACKNOWLEDGEMENTS

Agradeço primeiramente a Deus, pela saúde, lucidez e todas as inúmeras condições, muitas vezes invisíveis, que me foram concedidas ao longo do doutorado. Parte deste agradecimento nasce do receio de que eu deixe a graça passar despercebida, como nos alerta Santo Agostinho.

À minha família, meu alicerce incondicional, expresso minha profunda gratidão por todo o suporte, paciência e compreensão diante das exigências da vida acadêmica. Em especial, agradeço à minha esposa, Paola Fernandes, por sua presença constante, pelo companheirismo inabalável e pelo amor que sustentou os momentos mais desafiadores desta jornada.

Ao meu orientador, Professor José Antonio Apolinário Jr., agradeço por todos esses 11 anos (de 2015 a 2025) de inúmeras orientações em diversos trabalhos. Pela orientação firme e generosa, conduzida com a sabedoria de um professor experiente e a energia de um pesquisador apaixonado pelo que faz (assim como já havia feito durante minha trajetória no mestrado). Sua orientação constante foi determinante para a realização deste trabalho. Estendo meus agradecimentos ao meu co-orientador Professor José Manoel de Seixas, por suas valiosas contribuições, especialmente na definição e refinamento do escopo da tese no campo da estimação de DOA e localização. Agradeço ao meu co-orientador Professor Julio Cesar Duarte, cuja dedicação e atenção rigorosa foram fundamentais nessa reta final, especialmente na estruturação e inúmeras revisões desta tese. Agradeço ao Técnico Bonfim pelos serviços técnicos extremamente necessários para a consecução da fase experimental.

Agradeço à Marinha do Brasil pela oportunidade de desenvolvimento profissional e acadêmico, em especial ao Vice-Almirante Trovão, por seu apoio decisivo em momentos difíceis. Também registro minha sincera gratidão ao Vice-Almirante Zampieri, pela contínua confiança no meu trabalho. Agradeço ao meu orientador técnico na Diretoria de Sistemas de Armas da Marinha do Brasil, ETM Edilson Cezar Corrêa, pelo apoio constante e pela confiança depositada durante todo o período do doutoramento, assim como aos colegas de trabalho, cuja compreensão e incentivo foram fundamentais. Registro também meu agradecimento ao CC (AA) Valentim, pela colaboração contínua com as atividades do Setor de Capacitação e pela postura sempre cooperativa. De forma ampla, expresso minha gratidão a todos os amigos do ambiente profissional pelas palavras de apoio, pelas atitudes generosas e pela energia positiva que, em muitos momentos, serviram de ânimo para seguir em frente.

Rigel Procópio Fernandes

*“Tudo que é feito de carne é como se fosse mato,
e toda a glória dos homens é como se fossem flores campestres.
O mato seca, as flores caem.
Mas as palavras do Senhor Estarão sempre presentes.”
(Bíblia Sagrada, Pedro 24-25, 1)*

RESUMO

O constante desenvolvimento de veículos aéreos não tripulados (VANT) chama a atenção de setores como forças policiais, agricultura, serviços de resposta a emergências e defesa. Esse interesse crescente pode ser atribuído à sua simplicidade e custo-benefício. O fato de que esses dispositivos podem ser facilmente usados para atividades ilegais torna-se uma preocupação, já que poderia ser empregados para atividades como: terrorismo, vigilância não-autorizada e múltiplos tipos de cyber-ataques como *eavesdropping*, *jamming*, e *spoofing*. Portanto, a capacidade de estimar a localização e a direção de chegada de drones é necessária para assegurar a segurança de infraestruturas críticas e para lançar contra-ataques. O objetivo principal da tese é estimar a localização e a *direction of arrival* (DOA) do drone por serem os parâmetros mais importantes para o lançamento de contra-medidas usando acústica. As contribuições para a estimação da DOA de sinais acústicos emitidos por drones incluem: (1) o reconhecimento do potencial da condição zero cyclic sum (ZCS) para avaliar a consistência entre múltiplas estimativas de atraso temporal, explorando as várias combinações de picos primários e secundários da função de correlação cruzada; (2) a mitigação da complexidade computacional da busca exaustiva por meio de uma abordagem heurística baseada em algoritmos genéticos, mantendo a acurácia da estimativa; (3) a proposição de uma etapa de refinamento baseada em mínimos quadrados (LS), que complementa a condição ZCS e contribui para melhorar a robustez frente a picos espúrios; e (4) o desenvolvimento de um estimador de DOA passivo, de fácil implementação e sem necessidade de treinamento prévio, adequado para ambientes ruidosos. Os métodos propostos alcançaram taxas de acerto de até $94,0 \pm 3,1\%$ na identificação correta de DOA em experimentos com drones reais em ambiente reverberante. As contribuições para a localização de drones a partir de medidas de TDOA entre microfones incluem: (1) a formulação de métodos de otimização para explorar os picos primários e secundários das correlações cruzadas; (2) a validação dos métodos de otimização através de experimentos reais; (3) a comparação dos métodos com uma abordagem baseada em rede neural que demonstrou que validou a utilização dos estimadores propostos baseados em LS que atingem erros inferiores a 1 metro; e (4) a demonstração de que os métodos são flexíveis e podem ser utilizados tanto para estimação de DOA e quanto para localização. O menor erro médio de localização obtido com o método que explorou o ZCS e o LS (ZCS-LS) foi de $0,55 \pm 0,35$ m, enquanto o método baseado em rede neural atingiu $1,02 \pm 0,02$ m na melhor posição, confirmando a efetividade das abordagens propostas.

Palavras-chave: VANT; Drone; DOA; Sistemas inteligentes.

ABSTRACT

The constant development of unmanned aerial vehicles (UAVs) has drawn attention from sectors such as law enforcement, agriculture, emergency response services, and defense. This growing interest can be attributed to their simplicity and cost-effectiveness. However, the fact that these devices can be easily used for illegal activities becomes a concern, as they could be employed in actions such as terrorism, unauthorized surveillance, and multiple types of cyberattacks, including eavesdropping, jamming, and spoofing. Therefore, the ability to estimate the location and direction of arrival (DOA) of drones is necessary to ensure the security of critical infrastructure and to enable the deployment of countermeasures. The main goal of this thesis is to estimate the drone's location and DOA, as these are the most important parameters for launching acoustic-based countermeasures. The contributions to drone DOA estimation include: (1) the recognition of the potential of the zero cyclic sum (ZCS) condition for evaluating the consistency of multiple time delay estimates by exploring various combinations of primary and secondary peaks in the cross-correlation function; (2) the mitigation of the computational complexity of exhaustive search through a heuristic strategy based on genetic algorithms, while preserving estimation accuracy; (3) the proposal of a refinement step based on a least squares (LS) solution, which complements the ZCS condition and increases robustness against spurious peaks; and (4) the development of a passive DOA estimator that is simple to configure and does not require prior training, making it suitable for noisy environments. The proposed methods achieved success rates of up to $94.0\pm3.1\%$ in correctly identifying the DOA in experiments with real drones in reverberant settings. The contributions to drone localization based on TDOA measurements between microphones include: (1) the formulation of optimization methods that exploit both primary and secondary peaks of the cross-correlation functions; (2) the validation of these optimization methods through real-world experiments; (3) the comparison of these methods with a neural network-based approach, which confirmed the validity of the proposed LS-based estimators that achieved localization errors below 1 meter; and (4) the demonstration that the proposed methods are flexible and can be used for both DOA estimation and localization tasks. The lowest mean localization error obtained using the ZCS-LS method was 0.55 ± 0.35 m, while the neural network approach achieved 1.02 ± 0.02 m at its best position, confirming the effectiveness of the proposed techniques.

Keywords: UAV; Drone; DOA; Intelligent systems.

LIST OF FIGURES

Figure 1 – Drone shot down by Iraqi forces	15
Figure 2 – The two-layer feedforward NN employed in the initial experiments. Input normalization is not indicated in this figure.	36
Figure 3 – Acoustic drone signal emitted by a DJI Phantom 4	61
Figure 4 – Different time delay estimation problems	63
Figure 5 – GA search progress with fitness function and number of correct delays evolution	64
Figure 6 – Progress of GA search with fitness function	65
Figure 7 – Drone data acquisition set-up	66
Figure 8 – Spectrogram of drone noise collected from different distances	67
Figure 9 – The statistics of the cross-correlations $r_{x_i x_j}$	68
Figure 10 – Evolution ZCS and LS cost functions	68
Figure 11 – Histogram of the position of the correct set of delays (1,000 trials) . . .	69
Figure 12 – Comparison among LS, ZCS, and ZCS-LS DOA estimations (1,000 simulations)	69
Figure 13 – Cross-correlations of the acoustic signals collected from a Phantom 4 drone hovering in an outdoor environment	70
Figure 14 – Error of DOA according to the TDE additive noise	71
Figure 15 – Experimental results: drone DOA estimates 100 m away from the microphone	72
Figure 16 – Correct estimation of TDOA in a non-reverberating room (left) versus the effect of reverberation (right) on the RIR and the cross-correlation of the received signals.	74
Figure 17 – Spectrogram of the audio signal employed in the practical experiments.	75
Figure 18 – Classroom with eight microphones (MIC 1 to MIC 8) and three distinct speakers (SPKR 1 , SPKR 2 , and SPKR 3).	75
Figure 19 – Results for the position estimation.	77
Figure 20 – Histogram of the estimation errors.	78
Figure 21 – Set-up for the drone noise recording.	79
Figure 22 – Phantom IV drone spectrogram. (a) spectrogram with drone noise (b) background noise, i.e., all noises surrounding microphones except the drone noise.	79
Figure 23 – Geometry of TDOA-based localization methods.	80

Figure 24 – Results obtained using simulated delays (in number of samples) with additive noise, based on the dataset’s geometry and assuming the drone is hovering at position five. (a) Localization error as a function of the standard deviation of the additive noise. (b) Localization error when introducing a single outlier, while varying the noise standard deviation. (c) Localization error when introducing two outliers, while varying the noise standard deviation.	80
Figure 25 – Number of peaks used. (a) one pass GTS approach, and (b) GTS. . . .	81
Figure 26 – Peak extraction process. (a) cross-correlation from microphones 7 and 8 with the drone hovering in position 5, (b) the same cross-correlation with the peaks detected, i.e., a peak is the sample that has a higher amplitude than the two nearest neighbors, (c) simple peak extraction method (sample amplitude greater than the other two closest peaks), the true peak estimation is the 4th highest peak, (d) theoretical delay is the peak with the highest amplitude, (e) theoretical delay is closer to the 7th peak, and (f) theoretical delay is closer to the 9th peak. . . .	82
Figure 27 – Comparison of TDOA-based LS solutions using simulations.	83
Figure 28 – Simulations for the drone localization problem.	84
Figure 29 – Drone localization estimates. (a) conventional LS solution without taking into consideration τ_{max} (b) extended LS solution without taking into consideration τ_{max} (c) conventional LS solution with TDOA within τ_{max} (d) extended LS solution with TDOA within τ_{max}	85
Figure 30 – Localization results obtained using different number, n , of actual delays ES(n). (a) ES(5); (b) ES(6); (c) ES(25); and (d) ES(27).	87

LIST OF TABLES

Table 1 – All possible cyclic paths in a four-microphone array	26
Table 2 – All possible cyclic paths in a 7-microphone array	30
Table 3 – Search terms	39
Table 4 – Scientific and Technical Repositories	40
Table 5 – Inclusion criteria	40
Table 6 – Reasons for paper selection	53
Table 7 – Comparison of Related Works and Thesis Contributions	59
Table 8 – Quantitative results of this thesis and related works	59
Table 9 – Neural network regression results	76
Table 10 – Drone localization error and standard deviation	86

LIST OF ABBREVIATIONS AND ACRONYMS

AI	Artificial Intelligence
CNN	Convolutional Neural Network
DFT	Discrete Fourier Transform
DNN	Deep Neural Networks
DOA	Direction of Arrival
FFT	Fast Fourier Transform
GCC	Generalized Cross-Correlation
GEVD	Generalized Eigenvalue Decomposition
LS	Least-squares
MFCC	Mel Frequency Cepstrum Coefficient
ML	Maximum Likelihood
PHAT	Phase Transform
RIR	Room Impulse Response
SNR	Signal-to-noise-ratio
SRP	Steered-Response Phase
SSL	Sound Source Localization
STFT	Short-time Fourier transform
SV	Steering Vector
SVM	Support Vector Machine
TCDMA	T-shaped Circular Distributed Microphone Array
TDE	Time Delay Estimation
TDOA	Time Difference of Arrival
UAV	Unmanned Aerial Vehicles
ZCS	Zero Cyclic Sum

CONTENTS

1	INTRODUCTION	15
1.1	MOTIVATION	17
1.2	PROBLEM STATEMENT	18
1.3	OBJECTIVES	18
1.4	METHODOLOGY	19
1.5	JUSTIFICATION	20
1.6	STRUCTURE OF THE THESIS	20
2	LOCALIZATION AND DOA ESTIMATION METHODS	21
2.1	GENERALIZED CROSS CORRELATION ALGORITHM	21
2.2	ZCS AND LS COST METHODS	23
2.3	DOA ESTIMATION METHODS	24
2.3.1	EXHAUSTIVE SEARCH USING ZCS-LS	24
2.3.2	HEURISTIC SEARCH USING GENETIC ALGORITHMS AND ZERO CYCLIC SUM	27
2.4	LOCALIZATION METHODS	30
2.4.1	TDOA-BASED LS LOCALIZATION TECHNIQUE	31
2.4.2	THE TDOA-BASED OPTIMIZATION METHODS	32
2.4.2.1	GREEDY TDOA SELECTION	33
2.4.2.2	EXHAUSTIVE SUBSET SEARCH	33
2.4.2.3	INTEGRATION WITH LOCALIZATION PIPELINE	34
2.4.3	FINGERPRINT-BASED NN LOCALIZATION TECHNIQUE	35
2.5	PARTIAL CONCLUSIONS	36
3	LITERATURE REVIEW	38
3.1	LITERATURE REVIEW PLANNING	38
3.1.1	REVIEW OBJECTIVES	38
3.1.2	RESEARCH QUESTIONS	38
3.1.3	SEARCH TERMS	39
3.1.4	REPOSITORIES OF SCIENTIFIC INFORMATION	39
3.1.5	SEARCH CONSTRAINTS	39
3.1.6	IDENTIFICATION AND SELECTION OF STUDIES	40
3.2	ACOUSTIC-BASED COUNTER-DRONE SYSTEMS	41
3.2.1	DRONE ACOUSTIC SIGNAL	41
3.2.2	DRONE DETECTION AND CLASSIFICATION	42
3.2.3	DISTANCE ESTIMATION	44
3.2.4	PAYLOAD ESTIMATION	44

3.3	LOCALIZATION AND DOA ESTIMATION	45
3.3.1	TIME-DELAY ESTIMATION APPROACHES FOR LOCALIZATION AND DOA .	45
3.3.2	BEAMFORMING TECHNIQUES	49
3.3.3	AI-ORIENTED LOCALIZATION AND DOA ESTIMATION	51
3.4	POSITIONING WITHIN THE STATE OF THE ART	53
3.4.1	SUMMARY OF THE SELECTED WORKS AND DISCUSSION	53
3.5	PARTIAL CONCLUSIONS	58
4	DOA ESTIMATION RESULTS	60
4.1	GA-ZCS	60
4.1.1	PROBLEM STATEMENT AND ASSUMPTIONS	60
4.1.2	DRONE ACOUSTIC SIGNAL	61
4.1.3	TDE PROBLEMS WITH SIGNALS COLLECTED WITH UMA-8	61
4.1.4	SIMULATION RESULTS	62
4.1.5	EXPERIMENTAL RESULTS	63
4.2	EXHAUSTIVE SEARCH WITH ZCS-LS	65
4.2.1	DATA ACQUISITION	65
4.2.2	DRONE NOISE COLLECTED WITH SPHERICAL ARRAY	66
4.2.3	EFFECTS OF SIGNAL WINDOW LENGTH	66
4.2.4	DOA ESTIMATION WITH SIMULATED DATA	66
4.2.5	DOA ESTIMATION WITH EXPERIMENTAL DATA	68
5	LOCALIZATION RESULTS	73
5.1	THE REVERBERATION EFFECT	73
5.2	LOCALIZATION EXPERIMENTS	74
5.2.1	FIXED SYNTHETIC ACOUSTIC EMITTER EXPERIMENT	74
5.2.2	DRONE LOCALIZATION EXPERIMENT	78
5.2.3	COMPARISON OF THE LOCALIZATION TECHNIQUES	84
5.3	PARTIAL CONCLUSIONS	86
6	CONCLUSIONS	88
	BIBLIOGRAPHY	90
	APPENDIX A – PUBLISHED WORKS	102

1 INTRODUCTION

Unmanned aerial vehicles (UAVs) [1], also referred to as drones, have gained immense popularity in civil [2], industrial, and military applications [3]. These devices can perform a myriad of tasks, e.g., delivery [4], surveillance [5], mapping [6], photography [7], and agriculture monitoring [8]. In recent years, small UAVs have undergone significant improvements, including increased range, speed, and payload capacity [9].

Suojanen et al. [10] advocates that the dual-use capabilities of small UAVs enable them to be utilized by various actors for different purposes. Defense forces can use small UAVs to gain a strategic advantage over their enemies by conducting surveillance and reconnaissance missions or launching strikes against them, according to the works proposed by Watling & Waters; Kreps & Lushenko [11, 12]. On the other hand, terrorists can easily adopt low-cost drones to carry out attacks on targets, as stated by Pledger; Watling & Waters [9, 11]. Furthermore, individuals with malicious intent can use them to carry out criminal acts. It is, therefore, important for both public and private sectors to develop and implement effective countermeasures against the malicious use of small UAVs, as presented by Kunertova [3]. Figure 1 denotes a weaponized drone shot down by Iraqi forces, which was carrying a bomb as its additional payload, according to Watling & Waters [11].



Figure 1 – Drone shot down by Iraqi forces. Source: [11].

A counter-drone system comprises two stages, named threat evaluation and weapon assignment [13]. A significant challenge is the drone threat evaluation step, i.e., drone detection [14, 15, 16] and parameter estimation, such as DOA [17, 18, 19], localization [20,

21], model [22], and additional payload weight [23].

A drone threat evaluation system can detect and estimate drone parameters using different signals. For instance, radar [24], radio frequency [25], optical [26, 27, 28, 29], and acoustic signals [22, 30, 31] can be used. Other works employ two or more sensors to detect drones, e.g., optical and acoustic [32], acoustic and RF [33], video and acoustics collected from a drone [34]. Extensive literature about the detection and the parameter estimation of drones can be found in [35, 36, 37].

Drones can perform autonomous missions and, in this case, they will not emit RF signals, but they continue to produce acoustic noises [38, 39, 40]. Camera-based drone detection systems suffer from non-line-of-sight requirements, while radar-based counter-drone systems suffer from detecting drones due to the diverse materials in which the drone is built. Thus, it is possible to exploit their acoustic signature to detect and estimate the parameters of both kinds of drones, i.e., remotely piloted or autonomous drones. Conversely, acoustic signals are susceptible to attenuation and can be significantly affected by background noise, making it necessary to employ sophisticated techniques to detect drones and estimate their parameters, even in a noisy environment. These techniques must extract important features related to drone noise to accurately detect the presence of these possible threats. Acoustic techniques allow the estimation of drone parameters, e.g., localization, DOA, payload weight, or drone model. Therefore, the acoustic signal emitted by the drone is an important signature to accurately evaluate its parameters and allow a correct weapon assignment. Among these parameters, DOA estimation is required for the weapon assignment stage.

DOA estimation using acoustics has been extensively studied in the field of audio signal processing, and various methods have been proposed for sound source DOA estimation [41]. Several techniques have been proposed for DOA estimation of drones using acoustics, including beamforming, time-difference-of-arrival (TDOA) estimation, or machine learning-based approaches [22]. These techniques aim to accurately estimate the DOA of drones by exploiting the spatial and temporal properties of the drone's acoustic signature.

For instance, there have been reports of drones being used in the ongoing conflict between Russia and Ukraine [42]. These drones have been used by both sides for various purposes, including surveillance, reconnaissance, and even as weapons. In particular, quadcopter drones have been extensively used for surveillance, reconnaissance, and grenade launch. This utilization has raised significant concerns, as drones can be difficult to detect and cause significant damage to targets. Even a single drone from a swarm that breaches defenses has the potential to damage critical assets, such as counter-battery radar systems, thereby compromising defensive capabilities and exposing further vulnerabilities. Furthermore, we are learning on the fly that, due to their affordability, commercial drones

are becoming a preferred choice for replacing damaged drones, rather than investing in larger and more expensive military drones with enhanced capabilities, especially in an expensive and long-term war [43].

Therefore, the use of quadcopter drones in the conflict has highlighted the need for effective detection and countermeasure systems to ensure the safety of civilians and military personnel. The development of reliable DOA estimation techniques for drones using acoustics can assist in detecting and localizing drones, while also being used in conjunction with other detection systems, such as radar and cameras, to provide a comprehensive detection and tracking solution.

1.1 Motivation

The increasing use of drones in a plethora of applications has raised concerns regarding their potential to cause harm, especially in sensitive areas such as airports, military organizations, and public events. Therefore, the need for accurate and reliable counter-drone systems has become an important issue in ensuring public security [44].

It is possible to note the importance of counter-drone technologies in mitigating threats posed by unauthorized or malicious drone activities [45]. These technologies play an important role in detecting, identifying, and neutralizing drones that could otherwise compromise safety, privacy, and even national security. With the rapid advancement of drone capabilities, including enhanced flight range, payload capacity, and autonomy, counter-drone systems must evolve to address increasingly sophisticated threats. The motivation for developing robust counter-drone solutions is not only rooted in preventing attacks but also in maintaining operational continuity in environments where the presence of drones could impact air traffic, surveillance operations, or critical infrastructure. As drones become more accessible and widespread, the urgency to innovate in this domain becomes even more pressing, demanding reliable systems that can operate effectively in complex, real-world scenarios. estimating

The utilization of acoustics is recent and promising in this scenario. The main idea is to benefit from the unique acoustic features to estimate important parameters such as drone localization, speed, and trajectory, which are required for effective counter-drone measures. Acoustic-based systems offer an alternative when traditional detection methods, such as radar or vision-based systems, face limitations in environments with visual obstructions or radio frequency absence.

The majority of works in this field focus solely on the primary peak of the cross-correlation in time delay estimation (TDE). However, they overlook the potential information embedded in the secondary peaks, which can provide additional insights about multipath reflections and environmental conditions. By considering these secondary peaks,

it may be possible to improve the accuracy of DOA estimation and enhance system robustness in complex acoustic environments. This approach opens new avenues for developing more precise and reliable drone detection technologies, especially in environments where reverberation and noise complicate traditional signal processing methods. The difficulties of detecting and estimating drone parameters are discussed extensively in works such as those by Yang et al.; Chang et al.; Sun et al.; Yang et al. [46, 47, 48, 49].

1.2 Problem statement

Traditional counter-drone systems have limitations such as line-of-sight requirements, parameter estimation (such as drone model and detection of additional payload), and vulnerability to jamming [50]. Acoustic-based systems can operate in non-line-of-sight conditions, making them a valuable complement to existing counter-drone systems by providing a solution for drone detection and parameter estimation. Furthermore, using acoustics to detect drones has the potential to be cost-effective and can be deployed in various settings. Therefore, developing acoustic-based anti-drone systems is important for enhancing drone detection capabilities and improving public safety.

In the age of drone warfare, the ability to detect and neutralize UAVs has become an essential aspect of military operations. Traditional detection systems such as radar and cameras may have limitations in detecting drones, particularly when they fly at low altitudes or within urban environments. Acoustic-based drone detection systems can provide an alternative solution to detect and track drones in such scenarios.

The problem addressed in this thesis focuses on the challenge of estimating drone localization and DOA in highly noisy and reverberant environments, where traditional acoustic-based estimation methods often fall short. These challenging environments, such as urban areas and complex outdoor settings, present significant obstacles due to noise, multipath propagation, and environmental clutter.

1.3 Objectives

The primary objective of this work is to propose a novel acoustic-based counter-drone system capable of estimating both the localization and DOA of a quadcopter drone by analyzing the acoustic signals produced by its propellers. To address the challenges posed by periodic signals and multipath propagation, the proposed approach employs the ZCS condition and the TDE to improve delay accuracy by incorporating both primary and secondary peaks from cross-correlations. Furthermore, this work explores the use of an exhaustive search strategy with a computationally efficient cost function to refine the selection of time delays and enhance localization performance, as introduced by Fernandes,

Apolinário Jr. & Seixas [51].

The specific objectives include the following:

- conceiving a simulation system for the initial assessment of the performance of TDE-based DOA and TDOA-based localization estimators. This simulator is a first step towards understanding the effects of the multiple peaks of the cross-correlations on the DOA estimation and also to evaluate the TDE challenges that arise from the incorrect position of the microphones and the lack of subsample delay estimation;
- exploring the capabilities of the use of microphones to estimate drone localization and DOA; and
- proposing a method to identify a set of delays (including both primary and secondary peaks of the cross-correlation functions) that minimize localization and DOA estimation errors.

1.4 Methodology

This study was conducted through a combination of simulation design, experimental analysis, and algorithm development, aimed at evaluating and optimizing acoustic-based techniques for drone localization and DOA estimation. First, to validate and obtain the necessary knowledge for the chosen problem, a literature review was performed about the subjects of the thesis:

- counter-drone systems;
- drone noise;
- cross-correlation;
- drone direction of arrival estimation using acoustics; and
- drone localization estimation using acoustics.

The research was conducted using both simulated and real-world signals to validate the proposals that address the gaps identified in the literature. The performances of the algorithms employed in this work were assessed quantitatively. The main goal was to design methods that could be used in real-world scenarios. The simulated signals were synthesized based on the analysis of actual drone noise signals collected by a microphone array. For each specific objective solution, evaluation methods are proposed to demonstrate that the problem can be solved or that the error can be minimized.

1.5 Justification

The advancement of dual-use technologies is a central theme within the National Defense Strategy [52], highlighting the need for reorganization and modernization of the Defense Forces to respond to emerging security challenges effectively. Among the three strategic sectors identified as critical to national defense (territorial surveillance and control, defense of air and water spaces, and safeguarding maritime communication lines), one pressing challenge is the growing threat posed by UAVs or drones. These threats require innovative solutions, particularly for localization and DOA estimation in environments where traditional radar or visual tracking methods may be ineffective.

The results of this work can benefit the surveillance of critical structures and support the development of national technologies. In the Defense Forces, the application of the knowledge acquired during this research can contribute to building a counter-drone system to localize and destroy drones. The results presented in this work offer insights into key questions, such as how to effectively employ acoustic signals to maximize the accuracy of DOA estimation in counter-drone systems, and what strategies yield the most reliable DOA estimates based on inter-microphone delays. Addressing these questions contributes to a deeper understanding of the role of acoustics as a practical and viable sensing modality in counter-drone applications, offering guidance for researchers and practitioners in the development of robust acoustic-based localization systems.

1.6 Structure of the thesis

The thesis begins with Chapter 2 that introduces the proposed method, emphasizing its novel contributions to acoustic-based localization and DOA estimation. Chapter 3 presents a literature review addressing counter-drone systems, localization strategies, and DOA estimation techniques. This chapter also positions this thesis with state-of-the-art methods. Chapter 4 presents the DOA estimation results obtained through both simulations and real-world experiments. Chapter 5 follows with the experimental findings related to localization performance. Lastly, Chapter 6 summarizes the main conclusions and outlines directions for future research.

2 LOCALIZATION AND DOA ESTIMATION METHODS

This chapter describes in detail the proposed methods to estimate localization and DOA in highly noisy environments using acoustics. It also covers how to deal with the challenges associated with estimating localization and DOA using the primary peak and the secondary peaks of the cross-correlations. Section 2.1 presents the GCC method to estimate DOA, providing a detailed description of the cross-correlation function due to its importance for the subsequent sections. Section 2.2 discusses the cost functions ZCS and LS. Section 2.3 presents optimization methods for DOA estimation, including an exhaustive search method to reduce delay estimation errors and a heuristic search based on genetic algorithms. Finally, Section 2.4 defines methods that solely rely on TDOA measures to estimate the localization of a source. The methods described in Section 2.3 and Section 2.4 represent the core contributions of this work in advancing acoustic-based drone localization and DOA estimation under adverse conditions. The implementations of the methods can be found in [53].

2.1 Generalized Cross Correlation algorithm

To estimate the direction of arrival using the GCC function, it is necessary to estimate the TDOA between pairs of microphones. The TDOA is obtained from the peak of the cross-correlation $r_{x_i x_j}(\tau)$, defined as follows [54]:

$$r_{x_i x_j}(\tau) = \mathbb{E}[x_i(k)x_j(k - \tau)], \quad (2.1)$$

where $\mathbb{E}[\cdot]$ denotes the expectation operator and τ is the lag between two given sensors, x_i and x_j . In practice, statistical knowledge of the signals is not available, and Equation (2.1) is usually replaced by its time average estimate, given by:

$$\hat{r}_{x_i x_j}(\tau) = \sum_{k=-\infty}^{\infty} x_i(k)x_j(k - \tau) = x_i(\tau) * x_j(-\tau), \quad (2.2)$$

where $*$ is the convolution operator.

By taking the discrete Fourier transform of $\hat{r}_{x_i x_j}(\tau)$ and assuming real-valued signals, the cross power spectrum density between $x_i(k)$ and $x_j(k)$ can be expressed as follows:

$$\hat{R}_{x_i x_j}(e^{j\omega}) = \mathcal{F}\{\hat{r}_{x_i x_j}(\tau)\} = \mathcal{F}\{x_i(\tau) * x_j(-\tau)\} = X_i(e^{j\omega})X_j(e^{-j\omega}) = X_i(e^{j\omega})X_j^*(e^{j\omega}). \quad (2.3)$$

The cross-correlation can then be computed using Equation (2.4):

$$\hat{r}_{x_i x_j}(\tau) = \mathcal{F}^{-1}\{\hat{R}_{x_i x_j}(e^{j\omega})\}. \quad (2.4)$$

Adding a frequency weighting function $\psi(\omega)$ in Equation (2.4), we have the GCC as presented in Equation (2.5):

$$\hat{r}_{x_i x_j}(\tau) = \mathcal{F}^{-1}\{\psi(\omega)\hat{R}_{x_i x_j}(e^{j\omega})\}, \quad (2.5)$$

where classical cross-correlation corresponds to $\psi(\omega) = 1, \forall \omega$. A popular weighting scheme employed by the GCC is the PHAT, known to have good performance in reverberating scenarios [54]. PHAT also tends to have a sharper peak than classical GCC, increasing the performance of the TDE. Its weighting function is given by:

$$\psi^{\text{PHAT}}(\omega) = \frac{1}{|X_i(e^{j\omega})X_j(e^{-j\omega})|}. \quad (2.6)$$

Finally, the TDE is obtained as follows:

$$\hat{\tau}_{ij} = \arg \max_{|\tau| \leq \tau_{\max}} |\hat{r}_{x_i x_j}^{\text{PHAT}}(\tau)|, \quad (2.7)$$

where this function corresponds to Equation (2.5) with the weighting function described in Equation (2.6) and τ_{\max} is the maximum delay possible (in the number of samples) between microphone i and j , which occurs when the DOA has the same direction as the vector that connects sensors i and j , expressed as follows:

$$\tau_{\max} = \frac{|\mathbf{p}_i - \mathbf{p}_j|f_s}{v_s}, \quad (2.8)$$

where \mathbf{p}_i and \mathbf{p}_j are the position vectors of sensors i and j , v_s is the speed of sound, and f_s is the sampling frequency. The TDEs using inverse Fourier transform (iFFT) provide delays as integer multiples of the sampling period; this leads to errors that are particularly relevant in small arrays (small time delays between sensors) and with low sampling frequency. To mitigate this source of errors, we can interpolate the GCC, allowing more accurate estimations of the TDOA. In this work, cubic interpolation was applied across all points within the range $-\tau_{\max}$ and τ_{\max} , ensuring that all possible delay values are covered. In a 3-D scenario, d_{ij} is expressed as $\Delta \mathbf{p}_{i,j}^T \mathbf{a}_{\text{DOA}}$, where the TDE (in samples) is given by:

$$\tau_{i,j} = \frac{f_s(\mathbf{p}_i - \mathbf{p}_j)^T \mathbf{a}_{\text{DOA}}}{v_s} = \frac{f_s \Delta \mathbf{p}_{i,j}^T \mathbf{a}_{\text{DOA}}}{v_s} = \Delta \bar{\mathbf{p}}_{i,j}^T \mathbf{a}_{\text{DOA}}, \quad (2.9)$$

where $\Delta\bar{\mathbf{p}}_{i,j} = f_s \Delta\mathbf{p}_{i,j} / v_s$.

Based on the estimated delay, as given in Equation (2.7), and the delay based on the unknown vector \mathbf{a}_{DOA} , Equation (2.9), the least squares cost function, is defined as:

$$\xi = \sum_{i,j} (\hat{\tau}_{i,j} - \Delta\mathbf{p}_{i,j}^T \mathbf{a}_{\text{DOA}})^2, \quad (2.10)$$

for all possible pairs, $N = M(M - 1)/2$ for the case of M microphones.

Minimizing the cost function with respect to \mathbf{a}_{DOA} , the following is true:

$$\mathbf{a}_{\text{DOA}} = \mathbf{R}^{-1} \mathbf{d}, \quad (2.11)$$

where $\mathbf{d} = \Delta\bar{\mathbf{p}}^T \boldsymbol{\tau}$, $\boldsymbol{\tau} = [\tau_{1,2} \ \tau_{1,3} \ \dots \ \tau_{1,M} \ \tau_{2,3} \ \dots \ \tau_{M-1,M}]^T$ and $\mathbf{R} = \Delta\bar{\mathbf{p}}^T \Delta\bar{\mathbf{p}}$, $\Delta\bar{\mathbf{p}}$ are assembled as follows:

$$\Delta\bar{\mathbf{p}} = [\Delta\bar{\mathbf{p}}_{1,2} \ \Delta\bar{\mathbf{p}}_{1,3} \ \dots \ \Delta\bar{\mathbf{p}}_{1,M} \ \Delta\bar{\mathbf{p}}_{2,3} \ \dots \ \Delta\bar{\mathbf{p}}_{M-1,M}]^T. \quad (2.12)$$

The solution provided by Equation (2.11) may not have a unit norm, which must be ensured through normalization. Only after normalization can the azimuth and zenith angles be accurately computed using trigonometric functions.

Equation (2.11) provides all three coordinates only when using a spatial array. If a planar array is used, ambiguity occurs, and matrix \mathbf{R} is singular. When all sensors are in a plane (xy -plane for instance), we must adapt the sensor positions (\mathbf{p}_i) to suppress the coordinate associated with the perpendicular axis, which in this case is z . This way, \mathbf{R} is non-singular and Equation (2.11) provides $\hat{\mathbf{a}}_{\text{DOA}}^{\text{incomplete}} = [a_x \ a_y]^T$. As the \mathbf{a}_{DOA} must be unitary and we assume that the source is located above or below the array, it is possible to estimate the DOA.

2.2 ZCS and LS cost methods

The ZCS condition is derived from a configuration in which the selected microphones form a closed loop. The theoretical time delays obtained by the spatial distance between each pair of microphones that forms a closed loop are an Abelian group. The sum of all elements of an Abelian group is known to be zero, regardless of the order in which they are added [55].

This criterion finds practical application in digital signal processing within an array of sensors [51]. The main reason is that the TDE process is often plagued by several factors that introduce complexities and errors, impeding the attainment of accurate DOA [56]. Low Signal-to-Noise Ratio (SNR) [57] constitutes one of the primary obstacles, as it weakens

the discernibility of the signal of interest amidst background noise, leading to challenges in pinpointing the exact arrival time delay. Furthermore, multipath propagation [58], a phenomenon where signals arrive at the microphones through multiple paths, exacerbates the issue by causing time delay variations. This results in the reception of multiple, altered versions of the same signal, complicating the accurate identification of the original signal's true arrival time. Additionally, errors in the measurement systems, including calibration inconsistencies (attitude and geometry of the array) [59] or hardware imperfections, further contribute to inaccuracies in time delay estimation, subsequently impacting the precision of DOA calculations. The cumulative impact of these factors on the acoustic signals leads to multiple peaks within the cross-correlation [51], resulting in misleading time delay estimations when we only consider the peak of the cross-correlations with the highest amplitude. Consequently, this multitude of peaks affects the accuracy of DOA estimation. Thus, the ZCS condition may be explored to assess if sets of peaks of cross-correlations are coherent.

The LS is an extra method that aims to enhance accuracy and efficiency in DOA estimation, ensuring that the selected delays contribute significantly to the DOA estimation. It consists of estimating DOA with a given time delay vector and analyzing the sum of the squared error of the time delay calculated according to the DOA and each original time delay, according to Equation (2.13):

$$\xi = \frac{1}{N} \sum_1^N (\hat{\tau}_{ij} - \tau_{ij})^2. \quad (2.13)$$

2.3 DOA estimation methods

This section describes two methods to optimize the DOA estimation using secondary peaks of the cross-correlations between pairs of microphones and the cost functions introduced in Section 2.2.

2.3.1 Exhaustive search using ZCS-LS

The ZCS-LS method unfolds in two distinctive stages. In the initial stage, we perform an exhaustive search, calculating the ZCS cost function for all combinations of TDE. From this computation, we identify a small set of candidate time delay vectors characterized by the lowest ZCS cost function values. Transitioning to the second stage, we select a small subset of vectors from this set, evaluated concerning the ZCS cost function. This selection process is guided by the LS cost function, ensuring that the chosen candidate time delay vector not only possesses one of the lowest ZCS cost functions among all possible time delay vectors but also aligns optimally with the LS solution. By systematically navigating through these stages, the ZCS-LS method reduces the complexity

of an accurate DOA estimation that takes into account primary and secondary peaks of the cross-correlations.

For an exhaustive search, careful consideration is given to the number of microphones in the array. For instance, considering a 7-microphone configuration, selecting a subset of four microphones may be preferred due to considerations of computational feasibility and efficiency. The restriction to four microphones allows a manageable number of microphone pairs, specifically $\binom{4}{2}$ or six possible pairs. For each of these pairs, the recorded signals are segmented into windows, and the cross-correlation function is applied to estimate the time delay. Unlike conventional approaches that rely solely on the primary peak, this method considers both primary and secondary peaks of the cross-correlation, resulting in C candidate time delays per microphone pair.

The method explores the entire solution space, denoted as S , which is the set of all possible time delay combinations (C^6) within a 4-microphone array, considering C time delay candidates for each of the six cross-correlations. This choice ensures a systematic evaluation of feasible delay combinations, considering a low number C of time delay candidates. For instance, if $C = 10$, it is possible to perform an exhaustive search as the total solution space is $10^6 = 1,000,000$ potential time delay combinations.

In contrast, the computational complexity grows exponentially with an increase in the number of microphones [51]. For instance, in a 7-microphone array, where $M = 7$, the size of the solution space, denoted as S with C candidate delays, is determined by the formula $S = C^N$, where $N = 21$ for $M = 7$ microphones. If $C = 2$, the solution space is $2^{21} = 2,097,152$. As C increases, the solution space expands rapidly. For instance, with $C = 3$, the number of possible solutions reaches over 10 billion (10,460,353,203), and with $C = 4$, it exceeds 4 trillion (4,398,046,511,104) distinct combinations. This exponential growth underscores the practical advantages of utilizing a microphone array with a reduced number of microphones (M), ensuring computational efficiency in exploring the solution space for optimal DOA estimation.

The solution space for a 4-microphone array poses constraints to the cost function used in the exhaustive search process. This is where the reduced complexity cost functions using ZCS excel. By applying a low complexity cost function, the exhaustive search can efficiently explore these expansive solution spaces and navigate toward the global optimal solution.

Each combination of time delays is evaluated based on the ZCS cost function that quantifies their proximity to a zero-sum. The closer to a zero-sum, the more coherent the time delays for a given DOA are. The Exhaustive Search-Zero Cyclic Sum (ES-ZCS) method can efficiently identify the correct delays from a multitude of incorrect delays, particularly in situations with low SNR. The C candidate delays for each cross-correlation function $r_{x_i x_j}$ are the elements of each row of data matrix \mathbf{V} denoted as $\{\tau_{ij,1} \ \tau_{ij,2} \ \dots \ \tau_{ij,C}\}$

ordered in descending order according to the amplitude. For $M = 4$, which implies in $N = 6$, matrix $\mathbf{V}_{N \times C}$ with all candidate delays, is defined as

$$\mathbf{V} = \begin{bmatrix} \tau_{12,1} & \tau_{12,2} & \tau_{12,3} & \dots & \tau_{12,C} \\ \tau_{13,1} & \tau_{13,2} & \tau_{13,3} & \dots & \tau_{13,C} \\ \tau_{14,1} & \tau_{14,2} & \tau_{14,3} & \dots & \tau_{14,C} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \tau_{34,1} & \tau_{34,2} & \tau_{34,3} & \dots & \tau_{34,C} \end{bmatrix}, \quad (2.14)$$

such that it is possible to combine all delays and create the $N \times 1$ vectors \mathbf{v}_c to perform an exhaustive search, $1 \leq c \leq C$. The combination of all vectors \mathbf{v}_c forms the solution space S .

The ZCS cost function [51] plays an important role in this method, assessing the sum of delays in defined subsets forming closed loops to minimize instances of inaccurate zero-sum outcomes. Through an examination of all potential subsets forming closed loops and the summation of their results, this method diminishes the probability of encountering a zero-sum outcome that lacks the correct delays.

To facilitate the computational calculation of the ZCS cost function, a method that involves the identification and enumeration of closed loops based on the number of delays was devised. More specifically, when employing a $M = 4$ microphone array, we find that, with 3 delays, there are 4 closed loops, and with 4 delays, we observe 1 closed loop. The total number of closed loops, denoted as L , is therefore $L = 4 + 1 = 5$. It is important to note that the delay τ_{31} , which closes the loop, can be determined by taking the negative value of τ_{13} . Similarly, τ_{43} can be expressed as $-\tau_{34}$, and, in general, any delay τ_{ji} that closes the loop can be written as $\tau_{ji} = -\tau_{ij}$. By utilizing this property, we can compute all possible delays once and then manipulate them to identify the correct value of τ that closes the loop. This approach saves computational resources by avoiding redundant calculations and facilitates the determination of the correct delay for loop closure. The complete listing of all possible closed loops for 3 and 4 delays can be found in Table 1.

Table 1 – All possible cyclic paths in a four-microphone array

# delays	Closed loops			
3	τ_{12}	τ_{23}	τ_{31}	
	τ_{12}	τ_{24}	τ_{41}	
	τ_{13}	τ_{34}	τ_{41}	
	τ_{23}	τ_{34}	τ_{42}	
4	τ_{12}	τ_{23}	τ_{34}	τ_{41}

We can create a matrix $\mathbf{D}_{L \times N}$ based on Table 1 that can sum all delays, expressed by:

$$\mathbf{D} = \begin{bmatrix} 1 & -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}; \quad (2.15)$$

where each element of the resulting vector $\mathbf{f} = \mathbf{D}\mathbf{v}$ is the sum of all subsets (closed loops) such that the ZCS cost function, denoted as f , is then calculated as

$$f = \mathbf{f}^T \mathbf{f} = \|\mathbf{f}\|^2. \quad (2.16)$$

This ZCS cost function captures the squared norm of the resulting vector, encompassing the contributions from all subsets and providing a measure of coherence among the time delays concerning an arbitrary DOA. The best Z vectors, determined by the ascending order of ZCS scores, are stored and form a collection of potential solutions to the problem, achieved with low computational effort. Each combination of time delays serves as a plausible solution to the DOA estimation problem represented as a column within matrix $\mathbf{P}_{N \times Z}$:

$$\mathbf{P} = \begin{bmatrix} \tau_{11} & \tau_{12} & \tau_{13} & \dots & \tau_{1Z} \\ \tau_{21} & \tau_{22} & \tau_{23} & \dots & \tau_{2Z} \\ \tau_{31} & \tau_{32} & \tau_{33} & \dots & \tau_{3Z} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \tau_{N1} & \tau_{N2} & \tau_{N3} & \dots & \tau_{NZ} \end{bmatrix}, \quad (2.17)$$

while matrix $\mathbf{P}_{N \times Z}$ presents a range of potential solutions, it is essential to note that the time delay vector with the lowest ZCS score may not always constitute the optimal combination for accurate DOA estimation. Consequently, we have introduced a second phase to further refine the ultimate selection from the pool of Z candidate vectors. This additional step aims to enhance the precision and reliability of the chosen solution, ensuring that the DOA estimation is not solely dependent on the ZCS score but takes into account an additional LS cost function for a more precise outcome.

The ZCS-LS method is detailed in Algorithm 1.

2.3.2 Heuristic search using genetic algorithms and zero cyclic sum

One of the key advantages of GAs is their ability to handle large solution spaces and navigate through complex landscapes of possibilities. Unlike traditional search methods, GAs do not rely on explicit problem domain knowledge or constraints. Instead, they explore the solution space by iteratively generating and evaluating a population of candidate solutions.

Algorithm 1 Exhaustive search using ZCS and LS (M=4)

```

Compute all  $C$  candidate delays for every  $r_{x_ix_j}$ ;
for  $k = 1 : N$  do
    Compute  $r_{x_ix_j}, ij = 12$  to  $34$ 
    Obtain  $C$  candidate delays (larger peaks of  $r_{x_ix_j}$ )
     $\mathbf{V}_{k,:} \leftarrow [\tau_{ij,1} \ \tau_{ij,2} \ \dots \ \tau_{ij,C}]$ 
end for
Create a combination of time delays and compute ZCS;
for  $k = 1 : S$  do
     $\mathbf{P}_{:,k} \leftarrow \text{map } \tau_{ij,i} \text{ in } \mathbf{V}_{k,:}$ 
     $f = \text{ZCS}(\mathbf{P}_{:,k})$  Equation (2.16)
     $\mathbf{P}_{N+1,k} = f$ 
end for
Compute LS cost function of the  $Z$  time delay vectors with lowest ZCS;
for  $k = 1 : Z$  do
     $\xi = \text{LS}(\mathbf{P}_{1:N,k})$  Equation (2.13)
     $\mathbf{P}_{N+2,k} = \xi$ 
end for
Choose the time delay vector with the lowest  $\xi$  (LS cost function).

```

The enormous solution spaces pose significant challenges for traditional search methods, as exhaustively evaluating each possible solution becomes computationally infeasible. This is where GAs excel, by employing a heuristic search approach, efficiently exploring these expansive solution spaces while navigating toward local or global optimal solutions without evaluating every possibility.

GAs employ the concept of individuals represented as chromosomes, where each chromosome encodes a potential solution to the problem. These solutions are evaluated based on the ZCS fitness function that quantifies their proximity to a zero-sum. Through the use of selection, crossover, and mutation operators, GAs promote the exchange and recombination of genetic material between individuals, mimicking the genetic diversity and variation found in natural evolution.

The C candidate delays for each cross-correlation function $r_{x_ix_j}$ are the elements of each row of data matrix \mathbf{V} denoted as $\{\tau_{ij,1} \ \tau_{ij,2} \ \dots \ \tau_{ij,C}\}$. For $M = 7$, which implies in $N = 21$, the matrix $\mathbf{V}_{N \times C}$ with all candidate delays, is defined as

$$\mathbf{V} = \begin{bmatrix} \tau_{12,1} & \tau_{12,2} & \tau_{12,3} & \dots & \tau_{12,C} \\ \tau_{13,1} & \tau_{13,2} & \tau_{13,3} & \dots & \tau_{13,C} \\ \tau_{14,1} & \tau_{14,2} & \tau_{14,3} & \dots & \tau_{14,C} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \tau_{67,1} & \tau_{67,2} & \tau_{67,3} & \dots & \tau_{67,C} \end{bmatrix}. \quad (2.18)$$

The population, consisting of P individuals (or chromosomes), represents the collection of potential solutions to the problem. Each chromosome consists of genes,

denoted as g , which can take values from 1 to C , according to the number of candidate delays C . These genes allow for the exploration of all possible delay candidates in the matrix $\mathbf{V}_{N \times C}$. The chromosome is represented as a column of the matrix $\mathbf{P}_{N \times P}$:

$$\mathbf{P} = \begin{bmatrix} g_{11} & g_{12} & g_{13} & \dots & g_{1P} \\ g_{21} & g_{22} & g_{23} & \dots & g_{2P} \\ g_{31} & g_{32} & g_{33} & \dots & g_{3P} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ g_{N1} & g_{N2} & g_{N3} & \dots & g_{NP} \end{bmatrix}, \quad (2.19)$$

the matrix $\mathbf{P}_{N \times P}$ is a set of possible solutions, thus we create a vector of delays using each column of $\mathbf{P}_{N \times P}$, for instance, if $\mathbf{P}_{(:,1)} = \{1, 3, 2, \dots, 9\}$ the corresponding vector of delays, \mathbf{v} , corresponds to

$$\mathbf{v} = [\tau_{12,1} \ \tau_{13,3} \ \tau_{14,2} \ \dots \ \tau_{67,9}]^T.$$

The fitness function is an important component within the method, with the ZCS being employed for this purpose. This function evaluates the sum of delays within specific subsets that form closed loops, aiming to minimize the occurrence of erroneous zero-sum outcomes. By considering all possible subsets that create closed loops and summing their results, we reduce the likelihood of encountering a zero-sum result without the correct delays.

To better understand the contribution of the GA method, we analyze its estimations using an array of $M = 7$ microphones. This configuration renders Exhaustive Search computationally prohibitive. We find that, with 3 delays, there are 35 closed loops. Similarly, with 4 delays, we observe 35 closed loops. Moving on to 5 delays, we encounter 21 closed loops, while 6 delays give rise to 7 closed loops. Finally, when utilizing 7 delays, a single closed loop is formed. The total number of closed loops, denoted as L , corresponds to 99 in this context; $L = 35 + 35 + 21 + 7 + 1 = 99$. For instance, Table 2 denotes in the first line one closed loop, it is important to note that the delay τ_{31} , which closes the loop, can be determined by taking the negative value of τ_{13} , i.e., $\tau_{12} + \tau_{23} - \tau_{13}$. By utilizing this property, we can compute all possible delays once and then manipulate them to identify the correct value of τ that closes the loop. This approach saves computational resources by avoiding redundant calculations and facilitates the determination of the correct delay for loop closure. The complete listing of all possible closed loops for 3, 4, 5, 6, and 7 delays can be found in Table 2.

Table 2 – All possible cyclic paths in a 7-microphone array

# delays	Closed loops						
3	τ_{12}	τ_{23}	τ_{31}				
	\vdots	\vdots	\vdots				
	τ_{56}	τ_{67}	τ_{75}				
4	τ_{12}	τ_{23}	τ_{34}	τ_{41}			
	\vdots	\vdots	\vdots	\vdots			
	τ_{45}	τ_{56}	τ_{67}	τ_{74}			
5	τ_{12}	τ_{23}	τ_{34}	τ_{45}	τ_{51}		
	\vdots	\vdots	\vdots	\vdots	\vdots		
	τ_{34}	τ_{45}	τ_{56}	τ_{67}	τ_{73}		
6	τ_{12}	τ_{23}	τ_{34}	τ_{45}	τ_{56}	τ_{61}	
	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	
	τ_{23}	τ_{34}	τ_{45}	τ_{56}	τ_{67}	τ_{72}	
7	τ_{12}	τ_{23}	τ_{34}	τ_{45}	τ_{56}	τ_{67}	τ_{71}

We can, then, create a matrix $\mathbf{D}_{L \times N}$ based on Table 2 to sum all delays.

$$\mathbf{D} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & \dots & 0 & 1 \end{bmatrix}; \quad (2.20)$$

where each element of the resulting vector $\mathbf{f} = \mathbf{D}\mathbf{v}$ is the sum of all subsets such that the fitness function, denoted as f , is then calculated as

$$f = \mathbf{f}^T \mathbf{f} = \|\mathbf{f}\|^2.$$

This fitness function captures the squared norm of the resulting vector, encompassing the contributions from all subsets and providing a measure of the fitness or quality of the estimation. The GA-ZCS method is detailed in Algorithm 2.

2.4 Localization methods

This section presents two distinct approaches for sound source localization: a traditional method based on TDOA using LS optimization, and a data-driven approach using an NN model. The first approach applies a TDOA-based LS solution, which estimates the source position by using the TDOA between pairs of microphones. In addition, we present optimized TDOA methods that rely on primary and secondary peaks and the selection of subsets of TDOAs. The second approach uses an NN trained with features extracted from reverberation fingerprints of the environment.

Algorithm 2 Heuristic search using genetic algorithms with zero cyclic sum fitness function (GA-ZCS)

```

for  $k = 1 : N$  do
    Compute  $r_{x_i x_j}$ ,  $ij = 12$  to  $67$ 
    Obtain  $C$  candidate delays (larger peaks of  $r_{x_i x_j}$ )
     $\mathbf{V}_{k,:} \leftarrow [\tau_{ij,1} \ \tau_{ij,2} \ \dots \ \tau_{ij,C}]$ 
end for
Create population  $\mathbf{P}_{N \times P}$  of random integers  $[1, C]$ 
for  $k = 1 : N$  do
     $\mathbf{v}_i \leftarrow \text{map } \tau_{ij,x} \text{ in } \mathbf{V}_{k,:} \text{ according to } \mathbf{P}_{:,k}$ 
    First evaluation of individuals  $\mathbf{P}_{N \times P}$ 
end for
while  $f > 10^{-15}$  OR  $k < 2000$  do
     $\mathbf{v}_k \leftarrow \text{map } \tau_{ij,x} \text{ in } \mathbf{V}_{k,:} \text{ according to } \mathbf{P}_{:,k}$ 
    Crossover neighbor individuals
    Mutate
     $f \leftarrow \text{Evaluate individuals } \mathbf{P}_{N \times P}$ 
     $\mathbf{P}_{N \times P} \leftarrow \text{Select the best } P \text{ individuals}$ 
    Increment  $k$ 
end while

```

2.4.1 TDOA-based LS localization technique

The TDOA-based LS solution relies solely on the largest peak of the cross-correlations to estimate the source position. The conventional LS solution estimates the localization using Equation (2.21) [60]:

$$\hat{\mathbf{p}} = [\mathbf{I} \ \mathbf{0}](\mathbf{A}_1^T \mathbf{A}_1)^{-1} \mathbf{A}_1^T \mathbf{b}_1, \quad (2.21)$$

where \mathbf{A}_1 is defined as

$$\mathbf{A}_1 = \begin{bmatrix} (\mathbf{p}_2 - \mathbf{p}_1)^T & \Delta d_{21} \\ (\mathbf{p}_3 - \mathbf{p}_1)^T & \Delta d_{31} \\ \vdots & \vdots \\ (\mathbf{p}_M - \mathbf{p}_1)^T & \Delta d_{M1} \end{bmatrix}, \quad (2.22)$$

where \mathbf{p}_m is the position of the m^{th} microphone and the vector \mathbf{b}_1 is

$$\mathbf{b}_1 = \begin{bmatrix} b_{12} \\ b_{13} \\ \vdots \\ b_{1M} \end{bmatrix}, \quad (2.23)$$

with b_{1m} given by

$$b_{1m} = \frac{\|\mathbf{p}_m\|^2 - \|\mathbf{p}_1\|^2 - \Delta d_{m1}^2}{2}, \quad (2.24)$$

and Δd_{ij} calculated as

$$\Delta d_{ij} = \frac{v_s \tau_{ij}}{f_s}. \quad (2.25)$$

where v_s is the speed of sound and f_s is the sampling frequency.

The extended LS solution, which is also a TDOA-based solution, uses all the possible peaks of the cross-correlations. This solution can be achieved using the following closed-form equation [61]:

$$[\hat{\mathbf{p}}^T \hat{d}_1 \hat{d}_1 \dots \hat{d}_{M-1}] = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}, \quad (2.26)$$

where \mathbf{A} is defined, in this case, where distance measures from all cross-correlations are used, as

$$\mathbf{A} = \begin{bmatrix} (\mathbf{p}_2 - \mathbf{p}_1)^T & \Delta d_{21} & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ (\mathbf{p}_M - \mathbf{p}_1)^T & \Delta d_{M1} & 0 & 0 & \dots & 0 \\ (\mathbf{p}_3 - \mathbf{p}_2)^T & 0 & \Delta d_{32} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ (\mathbf{p}_M - \mathbf{p}_2)^T & 0 & \Delta d_{M2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ (\mathbf{p}_M - \mathbf{p}_{M-1})^T & 0 & 0 & \dots & \Delta d_{M(M-1)} \end{bmatrix}, \quad (2.27)$$

and the vector \mathbf{b} is

$$\mathbf{b}_1 = \begin{bmatrix} b_{12} & b_{13} \dots b_{1M} & b_{23} & b_{24} \dots b_{2M} \dots b_{(M-1)M} \end{bmatrix}. \quad (2.28)$$

The key element for estimating localization is the TDOA between microphones i and j , τ_{ij} . In ideal conditions, without noise, τ_{ij} can be estimated accurately, leading to a precise calculation of the source location.

However, in real-world scenarios, background noise can interfere with the correct estimation of τ_{ij} , resulting in a noisy estimate $\hat{\tau}_{ij}$. Additionally, in environments with strong reflections or reverberation, the true delay may not correspond to the maximum peak of the cross-correlation function, further complicating the localization accuracy. Therefore, there is a need for optimized methods to accurately estimate source localization based on TDOA measurements.

2.4.2 The TDOA-based optimization methods

To address inaccuracies inherent in noisy and reverberant environments, this thesis introduces and evaluates two algorithms that operate as TDOA selection and refinement strategies: Greedy TDOA Selection (GTS) and Exhaustive Search (ES). The core of these methods lies in selecting an enhanced subset of TDOA measurements that most likely correspond to the true acoustic delays between microphone pairs. This selection is guided by cost functions, introduced in Section 2.2, which compare the estimated TDOAs with the theoretical delays derived from the estimated location. These functions assess the

coherence of the time delays based on their consistency with possible solutions in the localization space.

The objective here is to minimize the average squared error between observed and theoretical TDOAs, or, in the case of ZCS, enforce geometric consistency via closed-loop delay relationships. Given an initial set of TDOAs, extracted from both primary and secondary peaks of the cross-correlation, the proposed algorithms apply distinct strategies to filter out unreliable measurements. The resulting refined TDOA subset is expected to better align with the actual drone position, thereby improving localization accuracy in such complex acoustic environments.

2.4.2.1 Greedy TDOA Selection

The GTS algorithm iteratively evaluates the impact of removing individual TDOA measurements on the localization accuracy. Starting with the full set of $N = \binom{M}{2} = \frac{M(M-1)}{2}$ TDOAs, denoted by $\boldsymbol{\tau}$, the algorithm removes one TDOA at a time in a greedy fashion. After each removal, it recalculates the average squared error between the estimated TDOAs, $\hat{\tau}_{ij}$, and the theoretical delays, τ_{ij} , using the cost function defined in Equation (2.13). A TDOA is retained only if its removal does not lead to an improvement in the localization estimate, thus refining the set to include only the most consistent and informative values.

If the removal of a TDOA leads to a reduction in the cost function, that TDOA is discarded. This process continues iteratively and sequentially across all TDOAs until no further improvement is achieved by discarding additional delays. Although the GTS algorithm is computationally efficient, its heuristic nature may lead to suboptimal results. The GTS algorithm is similar to ILS [62]; the latter removes the TDOA with the largest LS cost function in every iteration, while the former removes a TDOA only if the LS cost function decreases (a more conservative strategy). The complete procedure is detailed in Algorithm 3.

A simplified variant of the GTS algorithm described in Algorithm 3 is the one-pass GTS, which performs a single iteration over the N available time delays. In this approach, each delay is individually evaluated and removed only if its exclusion results in a decrease in the LS cost function. This variant offers the advantage of reduced computational complexity, as it avoids the need to evaluate all delays in multiple iterations. Additionally, it explores a different trajectory across the LS cost surface, potentially uncovering alternative solutions that may not be reached by the standard iterative GTS procedure.

2.4.2.2 Exhaustive Subset Search

The Exhaustive Subset Search (ESS) algorithm systematically evaluates every possible combination of n TDOAs, $n < N$. For each candidate subset, the average squared error between the theoretical and observed delays is computed. The subset minimizing this

Algorithm 3 GTS using LS cost

Input: Full set of N TDOA estimations, $\hat{\tau}$
Output: Reduced TDOA set τ^*
 Compute initial mean LS cost: ξ
while cost decreases **do**
 $cost_improved \Leftarrow \text{false}$
 $best_cost \Leftarrow \xi$
 for each active TDOA index k where $m_k = 1$ **do**
 Remove temporarily TDOA number k from $\hat{\tau}$, forming a new subset τ'
 Compute LS cost function ξ' using τ'
 if $\xi' < \xi$ **then**
 $best_cost \Leftarrow \xi'$
 $best_index \Leftarrow k$
 $cost_improved \Leftarrow \text{true}$
 end if
 Revert $m_k \Leftarrow 1$
 end for
 if $cost_improved$ **then**
 Remove TDOA number k from $\hat{\tau}$, updating the $\hat{\tau}$ TDOA vector
 $\xi \Leftarrow best_cost$
 end if
end while
return τ^*

cost function is selected as the optimal configuration. Although computationally intensive, this brute-force approach serves as a benchmark to assess the performance of more efficient algorithms. Algorithm 4 describes the ESS procedure.

Algorithm 4 ESS

Input: Full set of N TDOAs $\hat{\tau}$ and theoretical delays τ , subset size n
Output: Optimal TDOA subset τ^* according to Equation (2.13) criterion
 Generate all combinations \mathcal{C} of n elements from $\{1, 2, \dots, N\}$
 Initialize minimum cost: $\xi_{\min} \Leftarrow \infty$
for each combination $\tau'_c \in \mathcal{C}$ **do**
 Compute cost ξ_c using Equation (2.13) over τ'_c
 if $\xi_c < \xi_{\min}$ **then**
 $\xi_{\min} \Leftarrow \xi_c$
 $\tau^* \Leftarrow \tau'_c$
 end if
end for
return τ^*

2.4.2.3 Integration with Localization Pipeline

To validate the effectiveness of the ZCS method in selecting meaningful subsets of TDOA measurements for the localization task, we integrated ZCS with two distinct localization strategies: GTS and ES. The objective is to verify whether the TDOA subsets

returned by ZCS contribute to more accurate localization results compared to using all available TDOAs.

The localization pipeline is organized as follows:

- **Step 1 – TDOA Subset Selection using ZCS:** ZCS was applied to the complete set of pairwise TDOA measurements to identify a subset that satisfies a predefined consistency criterion based on the ZCS condition. This step is designed to use the secondary peaks to reduce the ZCS and LS cost functions.
- **Step 2 – Position Estimation:** Two combined localization techniques can be independently applied to the selected TDOA subset:
 - **ZCS + GTS:** The Greedy TDOA Selection algorithm computes the source position with an enhanced vector obtained using ZCS.
 - **ZCS + ES:** The Exhaustive Subset Search computes the LS cost function from all subsets composed of n TDOAs.

2.4.3 Fingerprint-based NN localization technique

Building upon the strategy of using secondary peaks from the cross-correlation function, we adopt an approach inspired by Sousa & Thomä [63] on RF emitter localization [63, 64]. This method incorporates not only the highest peak but also a selection of additional prominent peaks. In our adaptation, for each microphone pair among the $M(M - 1)/2$ possible combinations, we extract a set of the most significant peaks, specifically, their amplitudes and corresponding delays, and build a feature vector for each position based on this information. The drawback of this approach is that one needs to make measurements and extract features from each point in the room.

Let $x_i(n)$ and $x_j(n)$ be the signals recorded by the i -th and j -th microphones, respectively. The expression we used for estimating the GCC-PHAT is given as

$$\hat{r}_{ij}(\tau) = \mathcal{F}^{-1} \left\{ \frac{X_i(e^{j\omega})X_j^*(e^{j\omega})}{|X_i(e^{j\omega})X_j(e^{j\omega})| + \sigma_{ij}/\eta} \right\}, \quad (2.29)$$

where $X_i(e^{j\omega}) = \mathcal{F}\{x_i(n)\}$ and $X_j(e^{j\omega}) = \mathcal{F}\{x_j(n)\}$. Equation (2.29) uses the discrete-time Fourier transform (DTFT) for convenience, but our approach uses the fast Fourier transform (FFT). Moreover, the regularization term σ_{ij}/η is used to avoid a division by zero, with σ_{ij} being the standard deviation of $|X_i(k)X_j(k)|$, and k the FFT index with $0 \leq k \leq N - 1$. The regularization parameter η plays an important role in practical implementations and is set according to the amount of available data, i.e., the size of the processing block. The larger the number of samples, the larger the value of η , which helps ensure less noisy and more reliable estimates.

From the top, p , peaks of GCC-PHAT, we collect amplitudes and delays from all possible N pairs of microphones among the M microphones available to form the input feature vector

$$\begin{aligned} \mathbf{f} &= [\tau_{1,1} \cdots \tau_{p,1} \ m_{1,1} \cdots m_{p,1} \cdots \tau_{1,N} \cdots \tau_{p,N} \ m_{1,N} \cdots m_{p,N}]^T \\ &= [f_1 \ \cdots \ f_K]^T, \end{aligned} \quad (2.30)$$

where $\tau_{M-1,M}$ and $m_{M-1,M}$ represent the delay and magnitude of the largest peak of $\hat{r}_{M-1,M}(\tau)$, the GCC-PHAT of the last pair of microphones ($M-1$ and M). In this case, the number of features is $K = 2 \times p \times N$. We discard peaks with delays corresponding to distances larger than the distance of each pair of microphones.

Figure 2 illustrates the use of L sigmoid neurons in the hidden layer and three linear output neurons. Assuming $f_0 = z_0 = 1$, we express the z -th output of the hidden layer and the output vector (estimated position) as

$$z_l = \frac{1}{1 + e^{-\bar{z}_l}}, \quad \bar{z}_l = \sum_{\kappa=0}^K w_{l\kappa}^h f_{\kappa}, \quad \text{and} \quad (2.31)$$

$$\begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{bmatrix} = \underbrace{\begin{bmatrix} w_{10}^o & w_{11}^o & \cdots & w_{1L}^o \\ w_{20}^o & w_{21}^o & \cdots & w_{2L}^o \\ w_{30}^o & w_{31}^o & \cdots & w_{3L}^o \end{bmatrix}}_{\mathbf{W}^o} \begin{bmatrix} z_0 \\ z_1 \\ \vdots \\ z_L \end{bmatrix} = \mathbf{W}^o \mathbf{z}, \quad (2.32)$$

where $\mathbf{z} = [z_0 \ z_1 \ \cdots \ z_L]^T$.

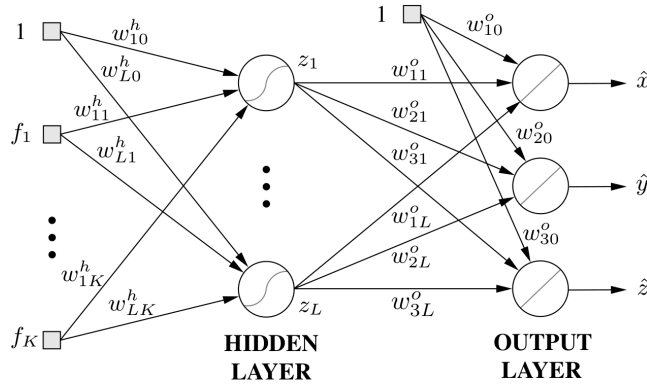


Figure 2 – The two-layer feedforward NN employed in the initial experiments. Input normalization is not indicated in this figure. Source: [65].

2.5 Partial Conclusions

This chapter presented the main methodologies employed and developed for drone localization and DOA estimation using acoustic signals in highly noisy environments. It

began with a detailed description of the Generalized Cross-Correlation (GCC) method, emphasizing its central role in TDOA estimation between microphone pairs, which is fundamental for both localization and DOA tasks.

Following this, we introduced two cost functions, i.e., ZCS and LS. These cost functions are designed to improve the reliability of TDOA estimation in scenarios where multiple correlation peaks may obscure the correct delay. These functions were integrated into both exhaustive and heuristic search frameworks, including a genetic algorithm approach, to enhance robustness and efficiency.

In the final part of the chapter, TDOA-based localization techniques were described, including both classical LS-based localization and learning-based models that leverage reverberation fingerprints. These approaches were developed to address limitations of traditional methods under real-world acoustic conditions.

The ZCS, LS, their integration into search strategies, and the neural network-based localization constitute important contributions of this work. They provide effective alternatives for accurate and robust acoustic-based localization and DOA estimation, particularly in highly noisy environments marked by noise, multipath, reflections, and reverberation.

3 LITERATURE REVIEW

This chapter presents the literature review on counter-drone systems, which aims to explain what drone parameters can be estimated using acoustics. This chapter is divided into four sections. Section 3.1 defines the review planning. Section 3.2 reviews the counter-drone systems. Section 3.3 reviews the specific tasks of localization and DOA estimation. Finally, Section 3.4 presents related works that are closely associated with the method developed herein. The main references used are presented by Khan et al.; Kadyrov et al.; AN et al. [66, 67, 68].

3.1 Literature review planning

This section outlines the methodology for conducting the literature review proposed by Faria et al. [69]. It details each step involved, including defining review objectives, formulating research questions, identifying search terms, selecting scientific information repositories, and establishing search constraints. Collectively, these steps form a framework that guides the selection of relevant works for this study.

3.1.1 Review objectives

This review serves the following purposes: (a) to comprehend counter-drone systems and drone parameter estimation; (b) to review techniques used to estimate the localization and DOA in the literature; (c) to review works that are closely related to the methods proposed in this thesis, i.e., to estimate the localization and DOA of drones using acoustics; (d) to evaluate the strengths, weaknesses, and gaps of existing methods; and (e) to establish a foundation for the proposed methods in this thesis by identifying opportunities for future research.

3.1.2 Research questions

The following research questions were formulated to guide the review: **(RQ1)** What are the primary methods and techniques used for acoustic-based drone localization and DOA estimation? **(RQ2)** How have these methods been adapted to perform in noisy and reverberant environments? **(RQ3)** What are the comparative advantages and limitations of these methods? **(RQ4)** What opportunities exist for future research to enhance performance in challenging environments?

3.1.3 Search Terms

The search strategy was developed to address the review research questions by identifying relevant works through a set of search terms. These terms focused on the key areas of acoustic-based drone localization and DOA estimation with applications to defense and law enforcement agencies. Table 3 presents the structured search terms used for identifying relevant literature in the review process. Table 3 organizes the search terms into five key dimensions: (1) **target**: which refers to the object of study, (2) **signal**: which is the type of signals used, (3) **sensor**: which is related to the sensors used to collect the acoustic signals, (4) **task**: which focuses on what activity is to be performed on the target using the signal and sensors mentioned, and (5) **measure**: which refers to the drone features extracted after signal processing.

Table 3 – Search terms

Dimensions	Terms
Target	“drone” OR “UAV” OR “unmanned aerial vehicle”
Signal	“acoustics” OR “acoustic”
Sensor	“microphone” OR “microphones”
Task	“DOA” OR “localization”
Measure	“time difference of arrival” OR “time delay estimation”

The search terms used in Section 3.2 focus on counter-drone systems; thus, the only dimensions used were “target” and “task”. The search terms used in Section 3.3 focus on localization and DOA estimation, so the dimensions used were signal, sensor, and task. As a result of the findings in Section 3.2 and Section 3.3, Section 3.4 uses all dimensions to perform a more restrictive search for works related to drone localization and DOA estimation of drones using acoustics. The search terms combination used in Section 3.4 to query academic repositories is: (“drone” OR “UAV” OR “unmanned aerial vehicle”) AND (“acoustics” OR “acoustic”) AND (“microphone” OR “microphones”) AND (“DOA” OR “localization”) AND (“time difference of arrival” OR “time delay estimation”).

3.1.4 Repositories of Scientific Information

Table 4 presents the academic repositories that gather relevant publications to the research, along with a brief description of each.

3.1.5 Search Constraints

The scope of this review was limited to papers published between January 2015 and September 2024, written in English, and relevant to drone DOA estimation and localization using acoustic signals. Only peer-reviewed journal articles and conference papers were included in this review. The literature review covers about ten years of

Table 4 – Scientific and Technical Repositories

Repository	Description
IEEE Xplore	Repository for articles on engineering research
ACM Digital Library	Focuses on computer science research
Elsevier	Multidisciplinary database covering a range of fields
JASA	Repository for articles on acoustics
Google Scholar	Indexes a wide variety of academic publications across multiple repositories

academic research and technical development. Papers selected to be examined from the repositories of scientific information in Section 3.4 are required to satisfy the criteria established in Table 5.

Table 5 – Inclusion criteria

ID	Criteria
IC ₁	The paper proposes or employs methods for drone localization or DOA estimation using the signals, sensors, and measures presented in Table 3 (aligned with RQ1)
IC ₂	The paper introduces or evaluates algorithms and techniques for estimating the DOA of drones in noisy or reverberant environments (aligned with RQ2)
IC ₃	The paper compares the performance of acoustic-based drone detection or DOA estimation methods, highlighting advantages and limitations (aligned with RQ3)
IC ₄	The paper discusses challenges, gaps, or trade-offs in using acoustic signals for drone detection and localization, especially in complex or noisy environments (aligned with RQ4)
IC ₅	The paper identifies opportunities for future research to improve acoustic-based methods for drone localization and DOA estimation (aligned with RQ4)

3.1.6 Identification and Selection of Studies

After conducting the searches and removing duplicate entries, 545 papers were selected. Google Scholar serves as a comprehensive search engine that crawls and indexes a vast range of academic repositories, journals, and conference proceedings. Its extensive coverage ensures that the selected papers from the repositories listed in Table 4 are included in its database. This coverage makes Google Scholar a reliable tool for academic research, providing access to significant results without requiring individual searches across multiple repositories. Thus, relying solely on Google Scholar is sufficient to gather relevant works from all the listed sources. Next, the titles, abstracts, and keywords were subjected to the inclusion criteria presented in Table 5. As a result, 490 papers were rejected and 55 accepted. Most of the rejected works focused on using drones to estimate the localization

and DOA of other targets or relied on different signals for the estimations, which is not aligned with the objectives of this thesis. The 55 papers that made it through the initial screening were subsequently subjected to a more detailed full-text review. Each paper was carefully evaluated based on the inclusion criteria outlined in Table 5. During this phase, 28 papers were further excluded due to not satisfying the inclusion criteria, leaving 27 papers that were considered highly relevant and selected for in-depth analysis. Of these, 15 papers were found to be highly relevant to the objectives and research questions of this thesis. These papers formed the basis for the theoretical foundation and identification of gaps in the literature.

The results of this review informed the proposed methodology, which aims to address existing challenges in drone localization and DOA estimation in complex environments.

3.2 Acoustic-based counter-drone systems

This section reviews the main characteristics of counter-drone systems, comprising the drone acoustic signal, drone detection and classification, distance estimation, and payload estimation. By examining these aspects, the section highlights the challenges faced and current advancements in counter-drone technologies.

3.2.1 Drone acoustic signal

Similarly to ships, drones emit noise mainly from their propellers, as well as from internal components like motors. Wang & Cavallaro [38] presents a detailed analysis of the acoustic emissions caused by the internal components of the drone. This was achieved by recording the drone functioning with no propellers. The acoustic profile of a UAV without propellers is rich in harmonic composition dominated by motor noise. The motor's fundamental frequency, or pitch, corresponds to the rising, stable, and dynamically varying motor rotation speeds. With four motors operating at different speeds, the superimposition of their pitches results in a nuanced and complex spectrum structure [38].

In the case of a normal UAV flight, i.e., with propellers, the noise signal exhibits two primary components: harmonic noise, characterized by energy peaks at isolated frequency bins, and broadband noise dispersing across the entire frequency band. Similar to a propeller-less UAV, where the drone operates without propellers, the pitch of the harmonic noise varies with motor rotation speed. The energy of the broadband noise is directly proportional to the propeller rotation speed, signifying that a faster rotation generates a more rapid airflow and, consequently, a more robust and distinct noise signature [38].

Wang & Cavallaro; Mukhutdinov et al. [39, 40] state that the acoustic signal emitted by a drone is caused mainly by the propeller, and it exhibits a time-frequency spectrum marked by distinct characteristics. The spectrum comprises narrowband harmonic noise

from the rotating motors and broadband noise from the propellers cutting through the air. Energy peaks at isolated frequency bins reveal a high correlation at harmonic frequencies, offering the potential for designing a time-frequency spatial filter to enhance sound from a specific direction.

3.2.2 Drone Detection and Classification

Rahman et al. [70] provides an in-depth exploration of various UAV detection and classification technologies, focusing on advancements in RF-based, visual data-based (images/video), acoustic/sound-based, and radar-based methods. Each method is examined, elucidating key challenges, proposed solutions, and future research directions. This review emphasizes the significance of RF-based UAV identification frameworks, highlighting advancements in signal processing techniques, machine learning algorithms, and multisensor fusion strategies to enhance detection accuracy and counter evasion tactics. Similarly, the utilization of computer vision techniques for visual-based UAV detection is discussed, emphasizing deep learning models, real-time processing capabilities, and robust algorithms to differentiate UAVs from other objects. Additionally, the study explores acoustic-based detection systems, focusing on sensor technologies, machine learning integration, and distributed sensor networks for improved UAV localization. Moreover, radar-based detection methods are explored, with an emphasis on radar system development, machine learning algorithms, and radar waveform diversity for enhanced detection performance. The study concludes with a discussion of the importance of data fusion techniques and highlights the potential of spectral and multispectral remote sensing imagery for precision UAV classification and detection.

Seidaliyeva et al. [71] reviews state-of-the-art techniques to detect and classify drones. The review examines various detection modalities, including radar-based, acoustic-based, RF-based, and visual-based approaches, while also addressing the inherent challenges posed by drones' dynamic behavior, diverse size and speed, and limited battery life. Integrating multiple sensory modalities is highlighted as a key strategy to enhance detection system robustness and accuracy. Early and late fusion techniques are discussed, alongside emerging approaches using Wi-Fi fingerprinting and cellular networks for effective drone detection. The main objective of this review is to guide future research endeavors and inform policymakers and practitioners in the field of UAV detection and classification.

The investigation into acoustic drone detection presented by Tejera-Berengue et al. [72] provides an understanding of the challenges and opportunities associated with detecting drones in real-world scenarios. By examining the impact of distance on sound propagation and its implications for drone detection, the study employs a range of machine learning techniques, from linear discriminant analysis to deep neural networks like YAMNet [73]. The evaluation, conducted using a carefully curated database and considering array signal

processing and ambient noise, demonstrates the efficacy of different training strategies in achieving effective detection at varying distances. Notably, the study reveals the potential for specialized detectors tailored to specific distance ranges, with advanced methods like YAMNet enabling detection up to 500 meters. Their findings underscore the importance of incorporating distance diversity into training datasets and highlight the suitability of deep-learning approaches for long-range drone detection. Future research directions include exploring advanced array processing techniques, enhancing data preprocessing methods, and investigating strategies to augment dataset size for training deep neural networks. Ultimately, the study suggests that acoustic drone detection holds promise for real-world applications, especially when integrated with other sensor modalities to provide comprehensive surveillance solutions.

Tejera-Berengue et al. [74] investigates the distance dependence of an acoustic signal-based UAV detection system. The study evaluates five detection methods by employing machine learning algorithms and a feature set encompassing Mel frequency cepstral coefficients (MFCC), pitch, spectral flux, among others. Linear discriminant analysis, multilayer perceptron, radial basis function network, support vector machine, and random forest are assessed for performance at various distances, revealing effective UAV detection with a performance decline as distance increases. The study concludes that acoustic detection is feasible at distances below 200 meters and could extend further in scenarios with more realistic interference conditions.

AN et al. [68] focuses on accurately estimating the total number of UAVs in a scene. Through the development of a UAV acoustic dataset featuring ten randomly flown combinations, acoustic information underwent preprocessing via time-frequency transformations to generate spectrogram images. These images were then input into a custom lightweight Convolutional Neural Network (CNN) model, achieving high test accuracy in estimating UAV numbers. The model's inference time performance on various edge computing devices was also assessed. According to AN et al. [68], future works aim to extend this approach to identify UAV models or types by incorporating information from additional sensors.

A recent review on auditory perception for unmanned aerial vehicles by Martinez-Carranza & Rascon [75] introduces a classification framework for UAV detection and classification into three categories: air-to-land, land-to-air, and air-to-air. This review highlights the dual roles of auditory perception in UAV systems—namely, detecting and classifying acoustic events, and localizing their sources. Additionally, Martinez-Carranza & Rascon [75] identifies the integration of microphones as a promising yet unresolved challenge for enhancing autonomous UAV navigation. An illustrative example is presented by Harvey & O'Young [76], who describe a non-cooperative collision-avoidance system that uses two microphones to estimate the detection range between aircraft.

Low signal-to-noise ratio (SNR) can be a problem in detecting and classifying sound events recorded from a drone, mainly because of ego-noise (the noise produced by the drone). This problem can be addressed using classical signal processing noise reduction algorithms, including frequency-spatial filtering techniques, effective in blind source separation problems proposed by Wang & Cavallaro [77]. More recently, methods based on Deep Neural Networks (DNN) described by Wang & Cavallaro [78] are also being used to enhance speech signals captured using drones. This work integrates single- and multi-channel DNN-based approaches for the enhancement of speech signals captured from drones.

3.2.3 Distance estimation

Kadyrov et al. [67] presents improvements in the acoustic drone distance estimation with a seven-microphone system, applying Steered-Response Phase Transform (SRP-PHAT) and narrowband frequencies for classification. Extensive tests, featuring drones like DJI Inspire 2 and Intel Falcon 8, showcased improved detection distances compared to the previous four-microphone system. The team recalibrated acoustic signatures to one meter, enabling the development of a straightforward method for estimating acoustic detection distances using the passive sonar equation. This advancement demonstrates the efficacy of the improved seven-microphone system in the drone detection system.

Ding et al. [79] introduces MUTES, a Multimodal Unmanned Aerial Vehicle 3D Trajectory Exposure System, responding to the rising demand for a comprehensive drone surveillance system. MUTES integrates a 64-channel microphone array for wide-range, high signal-to-noise ratio sound source estimation, coupled with long-range Light Detection and Ranging (LiDAR) and a telephoto camera for precise target localization. Implementing a coarse-to-fine, passive-to-active localization strategy, MUTES achieves semispherical surveillance with a broad detection range and high-precision 3D tracking. A dedicated environmental denoising model enhances fidelity by selectively isolating valid acoustic features from drone targets, overcoming traditional sound source localization challenges in noisy environments. Field experiments validate MUTES, demonstrating its farthest detection range, highest 3D position accuracy, robust anti-interference capabilities, and cost-effectiveness for identifying unverified drone intruders.

3.2.4 Payload estimation

Estimating drone payloads using acoustic signals has emerged as a significant area of research, driven by its potential applications in threat evaluation. This section synthesizes existing literature, focusing on techniques employing acoustic signals collected by either a single microphone or an array of microphones for payload estimation.

Ibrahim, Sciancalepore & Pietro [23] explores the novel concept of remotely detecting the payload weight of commercial drones through an analysis of their acoustic fingerprint. The research demonstrates that variations in motor speed and blade movement, influenced by payload changes, create distinct acoustic signatures using actual signals of a 3DR Solo drone. A single microphone was placed at a distance of 7 m from the hovering drone with 11 different payload weights. Utilizing MFCC components and Support Vector Machine (SVM) classifiers, the study achieves a 98% classification accuracy for payload detection in just 0.25 seconds.

Doster; Doster & Mullins [80, 81] pioneer the use of common cell phones for UAV acoustic payload detection, expanding prior studies conducted at close range. Addressing security concerns, it introduces the HurtzHunter prototype, demonstrating UAV payload detection with cell phones at distances ranging from 7 to 100 m using a single microphone. Using acoustic emissions, the system trains an SVM model with MFCC coefficients for payload classification, achieving accuracy ranging from 82.81% to 99.93%. This research not only extends the reach of UAV payload detection but also introduces an innovative approach using widely accessible cell phones for enhanced security in contested environments.

3.3 Localization and DOA estimation

This section provides an overview of the current state-of-the-art techniques for localization and DOA estimation.

3.3.1 Time-Delay Estimation approaches for localization and DOA

Borzino, Apolinário Jr. & Campos [56] discusses DOA estimation when the SNR is low. Although focusing on gunshot signals, the techniques employed there are also valid for other signals, since they combine the methods of generalized cross-correlation (GCC) with phase transform/maximum likelihood (PHAT/ML) proposed by Knapp & Carter [54], exhaustive search presented by Borzino, Apolinário Jr. & Campos [82], and the search for a fundamental loop proposed by Borzino, Apolinário Jr. & Campos [56]. This method searches for the best set of microphone pairs and makes a partial scan across the primary and secondary peaks of the cross-correlations, due to the computational efforts and the number of microphones used.

Firoozabadi et al. [83] explores multiple simultaneous sound source localization (SSL) in speech signal processing, a critical area of study. It navigates the balance between low computational complexity and high accuracy in SSL algorithms by combining a one-step-based method using generalized eigenvalue decomposition (GEVD) and a two-step method employing adaptive GCC-PHAT filters. This innovative amalgamation,

complemented by a unique T-shaped circular distributed microphone array (TCDMA), aims to enhance 3D multiple simultaneous SSL.

Bu, Zhao & Zhao [84] addresses the challenge of noise and reverberation in TDOA estimation. This research introduces two innovative methods aimed at effectively estimating TDOA in environments affected by noise and reverberation. The proposed methods leverage the linear phase structure observed across frequencies within a steering vector (SV) and capitalize on the absolute phases of SVs to mitigate potential noise and mathematical complications. By transforming the TDOA estimation into an optimization problem solvable via Newton’s method, the study presents experimental evaluations in simulated acoustic settings. In environments with moderate-to-high input SNR and low reverberation, their fast-search method demonstrates superior TDOA accuracy and computational efficiency.

Wang, Zhang & Wang [85] explores deep learning-based time-frequency masking to enhance TDOA estimation in challenging noisy and reverberant environments. Three novel algorithms are introduced to fortify conventional methods used for speaker localization, utilizing DNNs to identify cleaner time-frequency units for more accurate TDOA estimation. These algorithms exhibit robustness in scenarios with low SNR, high reverberation, and a low direction-to-reverberant energy ratio.

Liaquat et al. [86] reviews the use of microphone arrays for sound sensing, exploring the importance and limitations of ad-hoc microphones compared to other types. To address these limitations, the paper introduces specific approaches. Additionally, the study provides a detailed examination of existing methods for sound localization using microphone arrays, offering a comparative analysis and considering factors influencing the choice of one method over another. The aim is to establish a foundation for selecting the most suitable method for specific applications. A list of references on time delay estimation approaches can be found in the work proposed by Liaquat et al. [86].

Since its introduction, the GCC method proposed by Knapp & Carter [54] has captured significant attention in the academic community. Freire & Apolinário Jr.; Calderon & Apolinário Jr. [87, 88] focus on sniper detection, utilizing audio signals from gunshot recordings via a microphone array. Freire & Apolinário Jr. [87] employs the GCC-PHAT algorithm for DOA estimation, revealing that time-lags between the two largest peaks in the correlation functions align with muzzleblast and shockwave components. While the Phase Transform method excels in peak separation, the study concludes that muzzleblast DOA estimation based on the maximum correlation peak obtained by other GCC techniques is generally more accurate.

Freire [62] focuses on refining the DOA estimation in sniper detection through an exploration of TDE derived from cross-correlation functions. Employing an iterative least-squares (ILS) algorithm, the research identifies “matched lags,” minimizing errors in TDEs. In scenarios with low SNR, the ILS algorithm proves optimal, while the weighted least-

squares (WLS) [89] algorithm excels in high SNR conditions. WLS employs angle-related error propagation and the GCC-PHAT function quality, outperforming raw least-squares as verified in statistical analysis. The study introduces a methodology for DOA algorithm evaluation, recommending it for robust assessments. Particularly important for sniper localization, the proposed ILS algorithm excels in addressing the low SNR inherent in recorded gunshot audio, as supported by preliminary real-world data.

Borzino, Apolinário Jr. & Campos [82] addresses the critical task of gunshot DOA estimation, crucial for enhancing public and troop safety. The proposed algorithm is designed for scenarios with highly noisy signals, which commonly occur in sniper situations where the firing position is distant from the sensor array. In such scenarios, signal-to-noise ratio reduction poses a challenge to accurate DOA estimation. The paper introduces an innovative approach that combines an exhaustive search for optimal microphone pairs in the array, aiming for superior DOA estimation results and rapid response times across various shooting scenarios. The focus is particularly on highly corrupted signals where existing algorithms may falter. The proposed scheme's performance is evaluated using experimental data from both simulated and recorded gunshot signals.

Borzino, Apolinário Jr. & Campos [56] addresses the challenge of low SNR gunshot signals. The proposed algorithm combines the methods of exhaustive search (ES) and searching consistent fundamental loop (SCFL). The SCFL method utilizes the dominant peaks and secondary peaks of cross-correlation functions, often found in low SNR conditions. The ES-SCFL identifies the optimal set of microphone pairs and their correct cross-correlation function peaks using the ZCS condition [90], thereby enhancing the reliability of TDOA estimates and, consequently, DOA estimates. The algorithm's effectiveness is validated using real gunshot data recorded during a field experiment. The use of TDE presented good results in estimating gunshot DOA even in the presence of strong noise caused by drones [5, 91, 92, 93].

Fernandes, Apolinário Jr. & Seixas [51] addresses the challenge of accurately estimating the DOA of a drone in acoustically complex environments using a seven-microphone array. The focus is on improving TDE from a set of time delay candidates, particularly when dealing with strongly corrupted audio signals affected by noise and multipath. The traditional approach faces difficulties in accurately estimating TDE without relying on a line-of-sight assumption. The proposed solution utilizes genetic algorithms to perform a heuristic search for correct delays among possible pairs of microphones. A fitness function based on the concept of ZCS of closed loops is introduced, ensuring that the sum of theoretical delays in a closed loop equals zero. Experimental results, both in simulations and real-world trials, demonstrate the method's effectiveness in identifying correct delays, showcasing its potential for practical drone DOA estimation in challenging acoustic environments.

Shi et al. [18] implements a detection fusion algorithm and a TDOA estimation algorithm grounded in Bayesian filtering principles. This study employs two acoustic arrays, each comprising four microphones with a tetrahedron shape. The localization results are achieved using a closed-form LS solution. This work eliminates false peaks caused by multipath effects. The detection results are calculated every 0.5 s using an SVM model, and the results are demonstrated as the false alarm rate.

Jensen et al. [94] proposes innovative methods to mitigate the detrimental effects of reverberation on audio source localization. By incorporating models for both early reflections and the audio source itself, the authors introduce two iterative approaches for estimating the DOA of both the direct path and early reflections. The early reflections are effectively subtracted from the signal observations before localizing the direct path component, which reduces bias. Simulation results demonstrate the efficacy of these techniques, showcasing more accurate DOA estimation compared to state-of-the-art methods in both synthetic and real-world scenarios with reverberation.

Drémeau & Herzet [95] addresses the challenge of estimating the DOA of incident plane waves in scenarios where phase noise corrupts the received data (besides other additive noise). The proposed methodology adopts a Bayesian framework and employs a variational mean-field approximation to account for phase noise. By integrating sparse-enforcing distribution priors on DOA and Markov model priors on phase noise, the novel algorithm demonstrates superior performance compared to conventional beamforming and similar variational approaches with non-informative priors. Simulation results underscore the efficacy of the proposed approach in accurately estimating DOA, even in the presence of phase noise corruption.

Cui, Yu & Lu [96] introduces the constrained least squares (CLS) estimator, a novel approach for estimating the azimuth and elevation of a sound emitter in three-dimensional space using TDOA measurements obtained from an array of acoustic sensors. Addressing scenarios where the source emits transient signals, necessitating reliance solely on TDOA measurements for direction finding, the study highlights limitations of conventional linear least squares estimators due to inherent information loss during the linearization of nonlinear observation equations. To mitigate this issue, a constrained least squares estimator that employs both Lagrange multiplier and quadratic constraints to formulate the cost function is proposed. The resulting estimator offers an approximate closed-form solution, significantly reducing computational complexity while maintaining high accuracy. Theoretical analysis, supported by mathematical derivations, evaluates the estimator's performance in terms of mean square error. Through simulation and field experimental validation, the proposed method demonstrates superiority over traditional linear and nonlinear estimators, showcasing its potential for robust and efficient direction finding in practical applications.

Evers et al. [97] introduces a novel approach for distributed acoustic tracking by incorporating the coherent-to-diffuse ratio (CDR) as a measure of DOA reliability. Utilizing the CDR as the concentration parameter in the DOA-likelihood function, modeled by a von Mises distribution [98], enables tracking source positions over time at individual nodes using a von Mises filter. By evaluating the von Mises filter for a range of uninformative range hypotheses, the method employs network fusion to exploit spatial diversity among nodes, probabilistically triangulating the relevant source positions and range hypotheses. Realistic simulation results demonstrate significant improvements over classical approaches, enhancing accuracy by up to 39% compared to constant concentration parameter methods and up to 74% compared to least-squares source triangulation techniques [92].

Sewtz, Bodenmüller & Triebel [99] deals with the additive noise problem at varying degrees but not so well with a reverberation that arises naturally in indoor applications. Depending on the problem at hand, DOA [100, 101] and TDOA [102] can be estimated using a set of microphones conveniently distributed in space. Nonetheless, cross-correlation algorithms for TDOA estimation are sensitive to multipath propagation effects, resulting in inaccurate time difference estimates and severe position estimation errors. Therefore, most TDOA approaches aim to filter out multipath components to enhance performance. Early reflections play an important role in characterizing the acoustics of a room [103].

Ashraf, Hur & Park; Sousa & Thomä; Ribeiro et al.; Yapar et al. [104, 63, 105, 106] exploit the idea of generating multipath positioning fingerprint-based models to estimate the source location rather than discarding them. The use of neural networks taking advantage of reverberation fingerprint has been proposed by Yapar et al. [106] for urban localization and radio maps. As noted by Ribeiro et al. [105], the use of room impulse response to improve indoor localization does not provide good results, even under moderate reverberation conditions, while proposing the use of room reverberation modeling to improve indoor localization results under moderate reverberation conditions. However, under strong reverberation conditions, which is the case addressed, earlier reverberations can be stronger than direct path components, making the problem even harder to tackle using existing methods.

3.3.2 Beamforming techniques

Licitra et al. [107] provides a focused analysis of common beamforming algorithms, presenting both theoretical insights and recent applications in real cases. Rather than a broad exploration, the emphasis is on harmonizing the sector through a combined approach. The goal is to offer a resource for academics seeking theoretical understanding and technicians selecting algorithms for varied measurement conditions. With a lack of comparative studies in the literature, the authors address this gap, advocating for research in algorithm performance in similar scenarios. While acknowledging the limitations of

certain algorithms, the work generally recommends deconvolution algorithms (CLEAN-SC [108], DAMAS [109]) or MUSIC [110] for acoustic camera users due to their accuracy, even though they are slower and more complex. The authors propose a combination of algorithms for research purposes, anticipating future implementations in commercial acoustic camera software.

Ramos et al. [111] introduces the use of delay-and-sum to enhance sniper positioning estimates. The delay-and-sum beamforming is used for improved detection of shockwave and muzzle blast acoustic signatures. The approach not only enhances the signal-to-noise ratio, doubling the detection range for a 4-microphone array, but also demonstrates robustness in handling single- and multi-shot events and reflections, contributing to more reliable sniper location estimation. Other contributions to DOA estimation using delay-and-sum techniques can be found in the works proposed by Chiariotti, Martarelli & Castellini; Yang et al. [112, 113]

Lee, Hudson & Yao [114] investigates DOA for multiple acoustic sources using the approximate maximum likelihood (AML) algorithm. This algorithm facilitates the estimation of a DOA through an iterative search process, demonstrating versatility in both 2-D and 3-D scenarios. By employing blind beamforming techniques, the study shows the capability of the AML algorithm to estimate azimuth angles for sources in the far field of the array, as well as azimuth and elevation angles. The authors provide comprehensive analyses, including the calculation of the Cramér–Rao bound (CRB) [115] on DOA estimation, and introduce the concept of an isotropic array to enhance accuracy across the spatial domain. Simulation and experimental results validate the performance of the 3-D AML algorithm in scenarios involving multiple sources at varying azimuth and elevation angles.

Huang, Chen & Benesty [116] proposes an innovative approach to address the DOA estimation challenge within acoustic environments utilizing microphone arrays. The method initially transforms the received noisy speech signals into the STFT domain. Subsequently, a Householder transformation is constructed and applied to the multichannel STFT coefficients, segregating them into components dominated by the signal of interest and noise. By forming a cost function from the transformed coefficients, the method facilitates the extraction of DOA information by searching for extremum values within the angle range between 0 and 180 degrees. Simulation results presented in the paper demonstrate the effectiveness of this approach in achieving accurate DOA estimation.

Wajid, Kumar & Bahl [117] advocates an in-depth exploration of various algorithms—Bartlett Beamforming, Capons Beamforming, eigenvector, and Acoustic Intensity Vector for DOA estimation of both single and multiple sources employing an L-shaped Acoustic Vector Sensor (AVS). This specialized AVS configuration integrates three homogeneous sensors, each comprising omnidirectional microphones with a 14.14 mm aperture. To

facilitate experimental signal recording within the L-shaped AVS environment, the authors employ COMSOL Multiphysics, leveraging its Finite Element Method capabilities. Through systematic investigation and comparative analysis, the study offers valuable insights into the efficacy and performance nuances of different DOA estimation algorithms within the context of the L-shaped AVS configuration.

Kotus & Szwoch [118] proposes a calibration procedure for custom 3D AVS tailored for accurate DOA estimation. This calibration method addresses amplitude and phase differences among sensor components, crucial for precise DOA computation. Through experimental validation using low-cost MEMS microphones and DSP boards, the proposed procedure matches the accuracy of high-cost, factory-calibrated sensors. The study highlights the applicability of the calibration algorithm in practical scenarios such as environmental and traffic monitoring, offering a cost-effective solution for reliable sound source localization. Further research is suggested to expand the evaluation scope and refine the calibration approach for broader deployment.

Hu, Lu & Qiu [119] introduces a novel approach for multiple source DOA estimation using the maximum likelihood method in the spherical harmonic domain. By employing an efficient sequential iterative search of maxima on the cost function, the proposed method achieves superior performance compared to traditional beamformer-based and subspace-based methods. Notably, the method avoids the computational burden associated with high-dimensional grid search, making it suitable for both rigid-sphere and open-sphere configurations. Simulation and experimental validations conducted in various acoustic environments demonstrate the effectiveness and stability of the proposed method, highlighting its potential for practical applications in room geometry inference, source separation, and speech enhancement.

3.3.3 AI-oriented localization and DOA estimation

Genetic algorithms (GA) play an interesting role as a powerful heuristic search technique in solving complex problems [120]. It is inspired by the principles of natural selection and evolution, mimicking the process of survival of the fittest to find global or local optimal solutions. For instance, this heuristic search can be applied to optimize DOA estimation techniques according to Fernandes, Apolinário Jr. & Seixas [51].

Kassir et al. [121] reviews the cutting-edge applications of artificial intelligence (AI) in the domain of beamforming. Through an exploration of AI-centric beamforming studies, the work aims to elucidate and extract insights into the role of AI in enhancing beamforming performance. Beginning with an overview of beamforming and its adaptive algorithms, as well as DOA estimation methods, the analysis explores machine learning (ML) classes, neural network (NN) topologies, and efficient deep learning (DL) schemes. The paper further explores the optimal utilization of ML and NNs, both independently and

in conjunction with other applications such as ultrasound imaging, massive multiple-input multiple-output structures, and intelligent reflecting surfaces. Special emphasis is placed on the realization of beamforming or DOA estimation setups through DL topologies. Concluding with significant insights and a discussion on prospects and research challenges, the survey provides an overview of the evolving landscape of AI in beamforming.

Xiao et al. [122] introduces a novel high-resolution beamforming method employing genetic algorithms. By considering the sparsity of acoustic sources, the approach reconstructs the source vector through optimization within a sound propagation model. To enhance efficiency, the algorithm narrows down the search domain through prior correlation analysis. Numerical and experimental comparisons with conventional beamforming methods demonstrate the superior accuracy and robustness of the proposed genetic algorithm beamforming. Breaking through resolution limits, it accurately recovers the distribution of acoustic sources in two- and three-dimensional spaces. The work enriches existing high-resolution beamforming techniques, promising advancements in acoustic testing applications.

Kyritsis, Makri & Uzunoglu [17] presents a cost-effective small unmanned aerial system (UAS) acoustic detection system utilizing a four-microphone array that estimates DOA and UAS identification via machine learning techniques. Extensive outdoor experiments validate its efficacy in reliably detecting UAS at distances exceeding 70 m, offering enhanced situational awareness of the surrounding airspace.

Xiao et al. [123] advocates a learning-based approach for DOA from microphone array input, addressing limitations inherent in traditional signal processing methods like the classic least squares method. These conventional techniques are constrained by stringent assumptions on signal models and require precise estimations of TDOA, making them susceptible to noise and reverberation distortions. By contrast, the proposed learning-based approach uses a multilayer perceptron NN to learn from extensive simulated noisy and reverberant microphone array inputs, enabling robust DOA estimation. Extracting features from GCC vectors, the model effectively captures the nonlinear mapping to the DOA. Notably, the method's accuracy improves with the availability of more training data. Experimental evaluations on both simulated and real data demonstrate significant performance gains over the state-of-the-art LS method, with reduced root-mean-square error (RMSE) particularly evident in real-world scenarios such as meeting rooms.

Chakrabarty & Habets [124] advocates a novel CNN approach for broadband DOA estimation, wherein the phase component of short-time Fourier transform coefficients from microphone signals serves as direct input to the CNN. During training, the network autonomously learns the requisite features for accurate DOA estimation, and considering only the phase component of input facilitates training with synthesized noise signals, simplifying the dataset preparation compared to utilizing speech signals. Experimental

assessments validate the framework’s capability to generalize to speech sources and its robustness to noise, minor microphone position perturbations, and diverse acoustic conditions. Through both simulated and real data experiments, the study underscores the CNN’s adaptability and resilience, signaling the promising potential for practical DOA estimation applications.

3.4 Positioning Within the State of the Art

This section provides an overview of the works most closely related to the methods proposed in this thesis. It aims to establish the context and highlight key contributions from the literature that form the basis for the development of the approaches presented herein.

The subsection 3.1.6 outlines the execution process of the literature review, where a composite set of search terms was used to filter and select relevant papers from the repositories. The aim was to identify and evaluate academic works that align with the objectives and research questions set out for this thesis. subsection 3.4.1 presents a summary of the works selected in the previous section.

Table 6 – Reasons for paper selection

Reference	IC ₁	IC ₂	IC ₃	IC ₄	IC ₅
Blanchard, Thomas & Raoof [125]	✓		✓	✓	✓
Itare et al. [126]		✓		✓	
Sun et al. [127]		✓		✓	
Sedunov et al. [128]		✓		✓	
Sun et al. [48]		✓			
Chang et al. [47]		✓	✓		
Chervoniak et al. [129]	✓		✓		
Wu et al. [130]		✓		✓	
Wu et al. [21]		✓			✓
Lauzon et al. [131]	✓	✓			
Baggenstoss et al. [132]	✓	✓			
Chen, Yu & Yang [133]	✓			✓	
Faraji et al. [41]	✓				✓
Fernandes, Apolinário Jr. & Seixas [51]	✓	✓	✓		
Fernandes, Apolinário Jr. & Seixas [134]	✓	✓	✓	✓	

3.4.1 Summary of the selected works and discussion

The subsequent section provides a summary of the fifteen papers selected. Table 6 identifies the reasons for the selection of each paper.

Blanchard, Thomas & Raoof [125] considers the harmonic structure of UAV noise, employing a pitch detection algorithm and selective bandpass filtering to identify key

harmonics. While filtering within the antenna bandwidth reduces position errors, experimental results show that localization can still be accurate with only a few harmonics, with Kalman filtering applied to refine estimates.

Itare et al. [126] uses a microphone array with time domain Delay and Sum Beamforming, combined with a time-frequency representation of the beamformer's output, to focus on the UAV's acoustic signature for improved signal-to-noise ratio. By selecting specific frequency components, the method demonstrates robustness to noise and accuracy in localizing UAVs, even with limited spectral content or in the presence of multiple sources. The proposed method is highly effective, even in low-SNR environments (down to -16 dB), outperforming traditional delay and sum beamforming and other temporal filtering techniques. Real-world experiments with moving drones confirm the simulation results, with mean localization errors of less than 3° in azimuth and approximately 2° to 2.8° in elevation.

Sun et al. [127] puts forward a novel approach to indoor drone localization, addressing the challenges faced in GPS-denied environments, particularly in NLoS scenarios. Unlike existing methods that require significant hardware modifications or extensive environment instrumentation, Acoustic Inertial Measurement (AIM) uses the acoustic properties of drones for tracking and location estimation. The method uses a Kalman filter and Interquartile Range rule to reduce errors, achieving a 46% lower localization error compared to Ultra-WideBand-based (UWB-based) systems in complex indoor settings. In a $10\text{ m} \times 10\text{ m}$ scenario, it maintains sub-0.5 m accuracy when extended with distributed arrays. Additionally, AIM presents scalability by supporting spaces of arbitrary sizes through the deployment of distributed microphone arrays, offering accurate tracking in environments where traditional infrared systems fail.

Sedunov et al. [128] describes that the Stevens Institute of Technology developed the Drone Acoustic Detection System (DADS), an acoustic-based solution for UAV detection, tracking, and classification. Utilizing multiple microphone nodes arranged in tetrahedral configurations, DADS detects the DOA of UAVs based on propeller noise. The system has demonstrated effective real-time detection and tracking of UAVs such as the DJI Phantom 4, with a detection range of up to 350 meters and an average precision of 4 degrees (ranging from between 3.4° and 5.5°) depending on environmental conditions. Compared to more complex arrays, DADS proved to be a cost-effective and scalable solution. Future improvements aim to enhance classification capabilities, while a novel approach using simulated UAV sounds offers a low-cost testing alternative for acoustic system evaluations.

Sun et al. [48] offers a novel approach to indoor drone localization and tracking, particularly in GPS-denied and Non-Line of Sight (NLoS) environments. Unlike traditional methods that require extensive instrumentation or hardware modifications, AIM considers the acoustic signatures of drones, using a Kalman filter and the Interquartile Range rule

to reduce localization errors. The system, implemented with a standard microphone array, exhibits a 46% lower error rate compared to commercial UWB systems in complex indoor settings. In evaluations, AIM achieved average localization errors of 1.43 m in LoS, 1.89 m in partial LoS, and 2.08 m in full NLoS conditions. AIM's flexible design allows for scalable deployment with distributed microphone arrays, offering robust performance even in challenging environments.

Chang et al. [47] introduces a drone surveillance system using acoustic arrays for effective localization and tracking of UAVs. The authors propose a novel TDOA estimation algorithm, based on the Gauss *a priori* probability density function, to mitigate the challenges of multipath effects and low SNR. The system further employs a Kalman filter to track the drone's movement, utilizing TDOA results for localization. Field experiments validate the system's performance, demonstrating its accuracy and effectiveness in drone detection and tracking, even in challenging environments. The results demonstrated that over 95% of the localization errors were below 6 meters, and 80% were within 2 meters.

Chervoniak et al. [129] introduces an alternative method for passive detection and tracking of flying vehicles using specially developed hardware and software. The key contribution is a novel algorithm for time and frequency shift estimation based on resampling acoustic signals, which speeds up calculations compared to traditional methods. The system estimates TDOA and Doppler shifts to determine the position and velocity of flying vehicles. Experimental results show that the algorithm can detect small-sized vehicles at distances of several hundred meters, although performance improves significantly with larger aircraft. For small-sized drones, position estimation accuracy drops significantly beyond 100 meters, due to weaker acoustic signals and increased influence of environmental noise. The method offers a cost-effective solution for tracking flying vehicles and monitoring aircraft noise.

Wu et al. [130] addresses the challenges of UAV localization in low SNR environments by proposing a deep learning-based approach. Traditional acoustic methods often struggle with poor localization accuracy in such conditions. To overcome this, the authors present two key innovations: a MUSIC pseudo-spectral normalization technique to enhance DOA performance, and a time delay estimation neural network to improve DOA resolution in low SNR scenarios. Simulation results show that their approach achieves average localization errors below 0.5 m at 0 dB SNR and maintains accuracy within 1 m even at -10 dB SNR. Experimental results demonstrate that this method can localize UAVs within 20 meters, even with SNRs as low as -8 dB, showing its potential for real-world applications.

Wu et al. [21] addresses a novel drone localization system utilizing two tetrahedral acoustic arrays for effective localization within a 100-meter range. The system introduces a new TDOA estimation algorithm that combines GCC with probabilistic data association

to overcome challenges like multi-path effects and low signal-to-noise ratios. Additionally, a dynamic programming-based approach is proposed to optimize TDOA initialization. Extensive real-world experiments validate the system's performance, showing its effectiveness in providing real-time drone coordinates and enhancing the accuracy of existing TDOA estimation methods. Experimental results with a DJI Phantom 3 drone confirm the system's effectiveness: over 95% of both horizontal and vertical localization errors were below 6 m, even when the drone flew more than 100 meters from the arrays.

Lauzon et al. [131] explores the use of particle filtering for 3D sound source localization to detect and track rotary-wing unmanned aerial vehicles (RW-UAVs) using distributed microphone arrays. The method assumes that the dominant sound source is the RW-UAV, allowing for precise localization and tracking as long as its noise exceeds background noise. The qualitative results demonstrate effective 3D localization performance, although future work aims to enhance noise robustness by incorporating time-frequency masks and multi-source tracking. The authors also suggest the development of a large drone dataset for better motion state prediction and propose refining the system's precision and processing efficiency.

Baggenstoss et al. [132] addresses a phase-based acoustic detection and tracking algorithm for drones, focusing on efficiently tracking the TDOA of incoming signals from microphone pairs using DFT phase information. By forming solution curves corresponding to each TDOA and clustering their intersections, the algorithm determines the DOA without the need for computationally intensive grid searches, significantly enhancing efficiency compared to traditional beamforming methods. Simulations show that while the proposed algorithm is sub-optimal in terms of accuracy, it consistently loses no more than 2 degrees in DOA estimation compared to the maximum likelihood (ML) beamformer and approaches the Cramér-Rao lower bound (CRLB) at high SNR. In real-world scenarios, the phase-based algorithm successfully tracked drones through challenging maneuvers, indicating robustness against interference and the capability to generate phase-based spectrograms for sound classification. Overall, this approach not only improves computational efficiency but also maintains effective tracking performance in diverse conditions.

Chen, Yu & Yang [133] presents an innovative approach to accurately localize UAVs in challenging environments where radar and visual tracking are impractical. The proposed method employs an improved Empirical Mode Decomposition technique within an adaptive frequency window, allowing for effective smoothing and filtering of the UAV flight signals. Robust Empirical Mode Decomposition is used to decompose the signals into Intrinsic Mode Function components, facilitating detailed spectrum analysis. A sliding frequency window, optimized using a Grey Wolf Optimizer, is introduced to extract specific frequencies from the Intrinsic Mode Function, enhancing the separation of signal components. Subsequently, the Chan-Taylor localization algorithm, refined through weighted least squares, is applied

to calculate the target's position based on sensor time delay parameters. Validation through simulations and real-world signal tests demonstrates that the localization error remains below 5% within a $15\text{ m} \times 15\text{ m}$ measurement area, marking this method as an efficient and real-time solution for detecting small UAVs. Future work aims to address the challenges posed by multipath interference, particularly in environments where UAV signals are reflected and aliased.

Faraji et al. [41] addresses the challenging task of sound source localization, particularly in tracking flying objects, by introducing a novel fuzzy fusion and beamforming-based method using distributed sensor nodes. The proposed system comprises eight low-cost sensor nodes, each equipped with an array of synchronous MEMS microphones, designed to capture sound waves and record audio signals for offline evaluation. Each node estimates the direction of the sound source, which is then calibrated to account for installation errors, improving the accuracy of the location estimates. The calibrated directions are fuzzified and combined using fuzzy logic to achieve sound source localization. To validate the method, experiments were conducted with a quadcopter acting as a moving sound source in a wide outdoor environment measuring $240 \times 160 \times 80\text{ m}^3$. The algorithm achieved a mean distance error of 6.03 m compared to the quadcopter's GPS-derived trajectory, demonstrating effectiveness in real-time localization despite the challenges posed by the environment. The results indicate that the proposed approach balances high precision, robustness, and reasonable computational costs, making it a viable solution for practical applications in sound source localization.

Fernandes, Apolinário Jr. & Seixas [51] advocates an approach that improves the DOA estimation of drone noise using a microphone array, focusing on enhancing TDE in the presence of strong noise and multipath effects. Traditional methods often fail to accurately estimate TDE without a line of sight, particularly in complex acoustic environments where cross-correlation peaks may not correspond to true delays. To address this challenge, the authors employ genetic algorithms combined with a ZCS fitness function, which evaluates potential delays based on the principle that the sum of theoretical delays in a closed loop should equal zero. This method successfully identifies correct delays from a pool of candidates, including both primary and secondary delays, thereby significantly improving detection rates in experimental trials. The study provides compelling evidence of the effectiveness of genetic algorithms in resolving the TDE problem for drone signals by reducing the delay estimation error.

Fernandes, Apolinário Jr. & Seixas [134] addresses the challenge of accurate DOA estimation in noisy environments, particularly for applications in surveillance, security, and spatial audio processing. The authors propose a two-stage method that enhances DOA estimation by using secondary peaks of the cross-correlation function, which are often overlooked. In the first stage, a low-complexity cost function based on the ZCS condition

is employed to perform an exhaustive search of time delays between microphone pairs, encompassing both primary and secondary peaks. The second stage refines the estimation by applying an LS solution to a selected subset of the time delay combinations with the lowest ZCS cost. The method's effectiveness is demonstrated in the context of drone localization using a four-microphone array, achieving a notable accuracy of $94.0\% \pm 3.1\%$ when combined with the LS approach, compared to $89.4\% \pm 2.7\%$ using the ZCS method alone. Experimental results highlight the applicability of this refined DOA estimation approach in real-world scenarios, particularly by filtering out erroneous estimations that do not conform to the physical orientation of the microphone array.

3.5 Partial Conclusions

Based on the systematic literature review conducted in this chapter, two major research gaps were identified. First, most existing works rely on traditional TDOA estimation approaches that assume the main peak of the cross-correlation function corresponds to the true delay. This is often not the case in highly reverberant or noisy environments, leading to inaccurate DOA estimation. Second, although some studies attempt to incorporate learning-based methods or feature engineering, very few propose practical frameworks that combine signal processing with AI.

Motivated by these gaps, this thesis investigates the use of alternative cost functions capable of dealing with multiple candidate delays, which include primary and secondary peaks. In addition, this work presents the integration of heuristic search algorithms and neural networks to improve localization and DoA estimation accuracy in challenging acoustic scenarios.

Table 7 summarizes key characteristics of related works and highlights how the methods proposed in this thesis address the main limitations identified. The characteristics are: (1) if multipath is considered, (2) if the secondary peaks (Sec. Peaks) of cross-correlations are taken into account, (3) if optimization (Optim.) methods are employed, and (4) if AI-oriented methods are employed to estimate localization or DOA.

As shown in Table 7, the methods proposed in this dissertation directly address the shortcomings of traditional approaches by introducing innovative strategies that deal with ambiguous delay measurements and noisy environments. These strategies position this research as a novel and comprehensive contribution to the field of acoustic-based drone localization and DoA estimation.

Table 8 presents a comparative summary of the quantitative results reported by related works and the methods proposed in this thesis. The analysis highlights that the studies report results only on DOA or localization aspects. For instance, Itare et al. [126] achieves high DOA precision with azimuth errors under 3° and elevation errors between 2°

Table 7 – Comparison of Related Works and Thesis Contributions

Reference	Multipath	Sec. Peaks	Optim.	AI
Itare et al. [126]	✓	✓		
Sun et al. [127]	✓		✓	
Sedunov et al. [128]				
Chang et al. [47]	✓	✓		✓
Wu et al. [130]	✓	✓		✓
Wu et al. [21]	✓			✓
Lauzon et al. [131]			✓	
Baggenstoss et al. [132]	✓		✓	✓
Chen, Yu & Yang [133]	✓	✓		
Faraji et al. [41]				✓
Fernandes, Apolinário Jr. & Seixas [51]		✓	✓	✓
Fernandes, Apolinário Jr. & Seixas [134]	✓	✓	✓	
This thesis	✓	✓	✓	✓

and 2.8° , while Wu et al. [130] demonstrates robust localization with errors below 0.5 meters at 0 dB SNR. Some works, like Wu et al. [21] and Chang et al. [47], report localization accuracies with 95% of errors under 6 meters. In contrast, this thesis not only addresses both DOA and localization but also achieves state-of-the-art results in both metrics. The proposed method reaches a DOA accuracy of $94.0\% \pm 3.1\%$, and when integrated with the ZCS-LS approach, yields a localization error as low as 0.55 ± 0.35 meters. These results underscore the effectiveness and robustness of the proposed methods, especially under highly noisy and reverberant environments where many traditional approaches fail to provide consistent performance.

Table 8 – Quantitative results of this thesis and related works

Reference	DOA	Localization
[126]	$\phi < 2.9^\circ, \theta < 4.7^\circ$	N/A
[127]	N/A	< 0.5 m (10×10 m area)
[128]	$[3.4^\circ, 5.5^\circ]$	N/A
[47]	N/A	95% < 6 m, 80% < 2 m
[130]	N/A	< 1 m
[21]	N/A	95% < 6 m
[131]	N/A	Qualitative only
[132]	$< 2^\circ$	N/A
[133]	N/A	$< 5\%$ (15×15 m area)
[41]	N/A	Mean: 6.03 m
[51]	83.5% accuracy	N/A
[134]	$94.0\% < 5^\circ$	N/A
This thesis	$94.0\% < 5^\circ$	60% < 2.73 m, 40% < 1.07 m

4 DOA ESTIMATION RESULTS

In this chapter, we present the outcomes achieved through the utilization of the proposed drone DOA estimation methods in Section 2.3. We evaluated the method using simulations and also actual data to reveal the key findings. Section 4.1 presents the results related to the drone DOA estimation using a heuristic search scheme. Section 4.2 presents the results of acoustic-based drone DOA estimation using the Exhaustive Search.

4.1 GA-ZCS

This section presents the results related to the drone DOA estimation using genetic algorithms through experimentation with simulations and actual data.

4.1.1 Problem statement and assumptions

The signals emitted by the drone are characterized by a low SNR, indicating the presence of intense background noise and potential multipath effects. To conduct our simulations, we utilized the geometry of a compact acoustic array consisting of seven microphones, specifically the MiniDSP UMA-8 model [135]. Using this array, we calculate the theoretical delays of acoustic front waves emitted by the drone, considering both zenith and azimuth angles denoted as θ and ϕ , respectively. These theoretical delays represent the ideal estimates in the absence of background noise and multipath effects. Additionally, we acknowledge that estimated delays might manifest as secondary peaks within a cross-correlation analysis.

To simulate real-world conditions, we construct a data matrix denoted as $\mathbf{V}_{N \times C}$, according to Equation (2.18). Each row of the matrix contains a combination of theoretical delay and other random delays that can occur due to factors such as additional noise.

Moving on to the experimental phase, we gather two sets of acoustic drone signals, each lasting 20 seconds, employing the UMA-8 microphone array of $M = 7$ microphones. From these signals, we estimate potential delays and construct the data matrix $\mathbf{V}_{N \times C}$, where each row of this matrix contains candidate delays, i.e., peaks of the cross-correlations.

Consequently, our primary challenge lies in identifying the correct delays among various cross-correlation peaks arising from different microphone pairs. The simulations we conduct serve as a proof of concept for the proposed method, while the experimental trials provide evidence of the benefits this method offers in addressing the drone DOA estimation problem. Notably, the actual drone signals employed in this study were obtained during the hovering of a DJI Phantom 4 quadcopter.

4.1.2 Drone acoustic signal

Figure 3 provides a visual representation of the drone hovering signal, showcasing 10,000 samples in the time domain. Additionally, a spectrogram computed with a sample rate of 48 kHz is presented. It is worth noting that the drone noise is primarily concentrated in the frequency range below 5 kHz. However, under favorable conditions and when the drone is near the microphone array, it becomes possible to capture drone noise in higher frequencies, reaching up to 13.5 kHz.

For a more complete exploration of the acoustic characteristics of drone noise, readers may refer to the studies presented by Wang & Cavallaro; Wang & Cavallaro; Mukhutdinov et al. [38, 39, 40]. These references show the intricacies of drone noise analysis and provide insights into the subject matter.

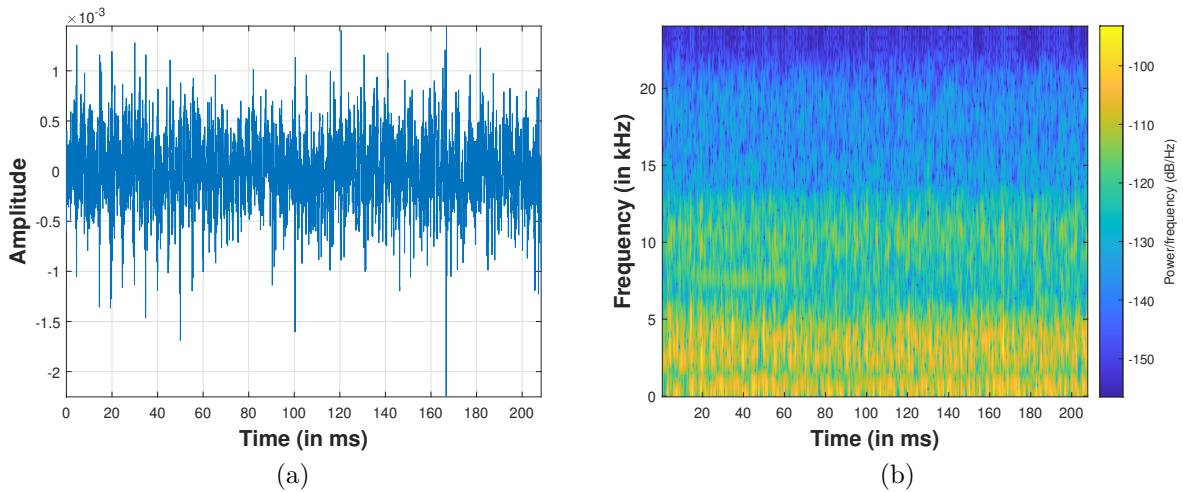


Figure 3 – Acoustic drone signal emitted by a DJI Phantom 4. (a) Time domain signal; and (b) Time-frequency representation.

4.1.3 TDE problems with signals collected with UMA-8

One common source of error in cross-correlation-based time delay estimation is the presence of noise. When the signals being correlated are contaminated by noise, it can introduce spurious correlations and lead to incorrect time delay estimates. The noise can distort the shape of the cross-correlation function, resulting in erroneous peak positions or false peaks that do not correspond to the true time delay.

Another factor that can cause inaccurate time delay estimation is the presence of reverberation or multipath in the signals. Reverberation can significantly affect the shape and amplitude of the cross-correlation function, making it difficult to accurately identify the true peak representing the time delay. The reflections and multiple paths of sound

propagation can create additional peaks or distort the main peak, leading to incorrect estimates.

Figure 4 illustrates several pertinent problems associated with TDE. In Figure 4 (a), we observe an accurate TDE both with and without interpolation, even in the presence of low levels of background noise. Figure 4 (b) showcases the benefits of interpolation, where a cross-correlation with fractional delays in samples yields a precise estimation. Figure 4 (c) presents a distorted function with a false main peak. Finally, Figure 4 (d) highlights a scenario with significant noise, wherein a secondary true peak emerges.

The abundance of different peaks of the cross-correlations involving drone noise instigates intriguing possibilities in the context of experimental trials. These possibilities include exploring secondary peaks using ES, employing the peaks of interpolation, utilizing classical cross-correlation peaks, and considering samples before and after the main peak. By considering these aspects, it is possible to enhance our understanding and refine the TDE methodology in practical scenarios.

4.1.4 Simulation results

The initial evaluation of the proposed approach for drone DOA estimation involved simulated delays, which approximate the potential delays based on the array geometry utilized herein. Figure 5 illustrates the evolution of the fitness function and the number of matching delays with the theoretical 21 delays. The ZCS fitness function serves as a guiding measure for the algorithm, leading to the enhancement of individuals within the genetic algorithm. By using this fitness function, the algorithm is directed towards improving the accuracy of delay estimates by identifying the correct peaks among the 10 possible delays for each cross-correlation function. This facilitates the overall improvement of the algorithm's performance in estimating the delays more effectively. However, it is important to note that GA may converge to a local minimum, and also that reaching the global minimum can be time-consuming.

Furthermore, the simulation results provide insights into the behavior and practical effectiveness of the GA when applied to the TDE problem under complex acoustic conditions. The nature of GA enables a broad exploration of the solution space, allowing it to find delay combinations that satisfy the ZCS condition, even when multiple cross-correlation peaks complicate the estimation process. As observed in Figure 5, the GA is capable of gradually improving the number of correctly identified delays over generations, guided by the fitness function. While convergence to a global minimum is the ideal scenario, in practice, even convergence to a local minimum can be sufficient. This is particularly true if the GA is combined with a second-stage optimization step that discards delays inconsistent with the geometric constraints of the array. In this hybrid approach, the GA identifies a promising subset of delays, from which the least consistent can then be

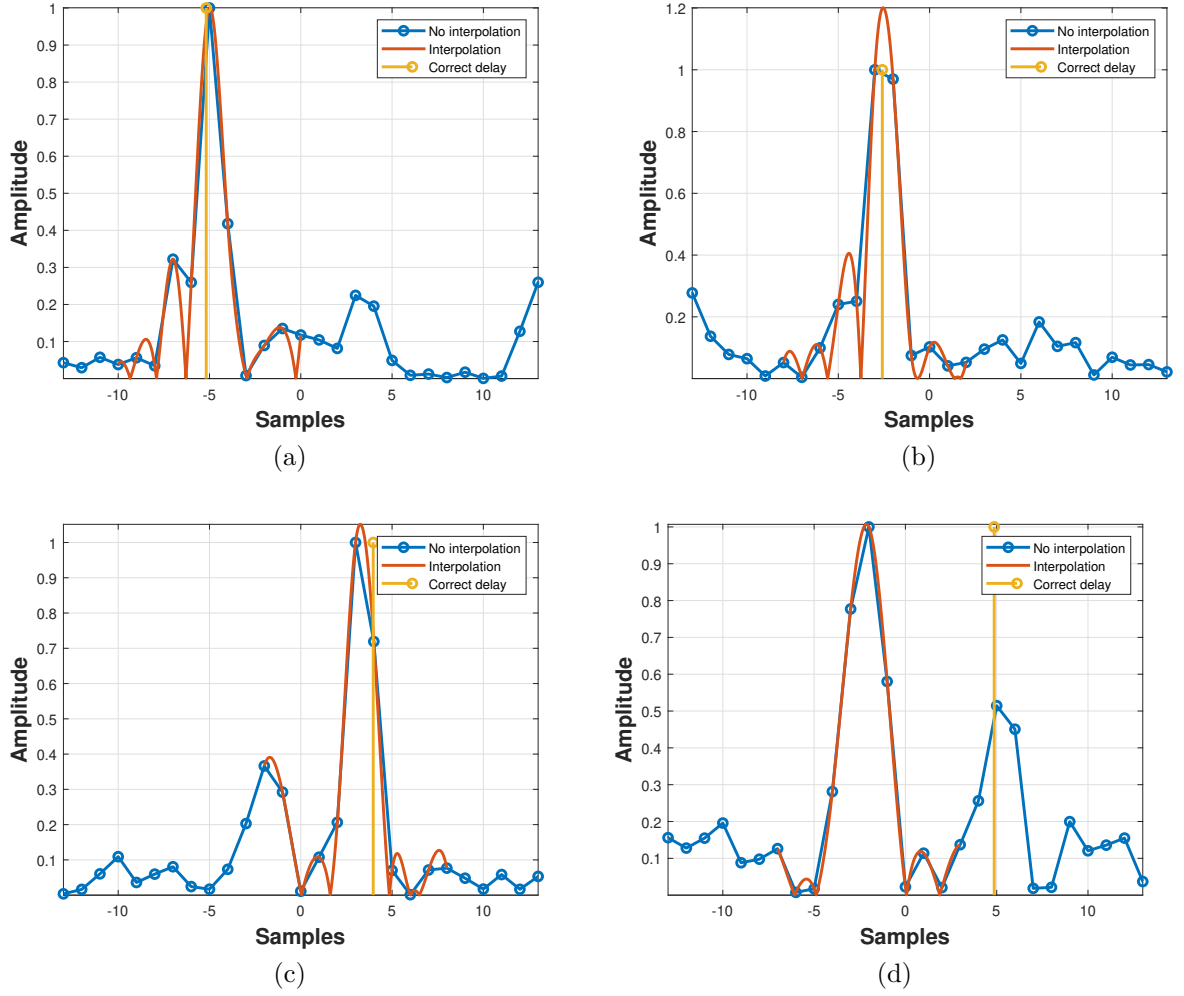


Figure 4 – Different time delay estimation problems. (a) Accurate time delay estimation with and without interpolation (b) More accurate time delay estimation with interpolation (c) Distorted function with a false peak and (d) Secondary true peak.

removed based on an LS cost or similar criterion. Once a few accurate TDEs are identified, they can anchor the solution and enable robust DOA estimation, even if the remaining delays include errors.

4.1.5 Experimental results

After conducting simulation tests, we proceeded to test the proposed method with actual drone noise signals. The signals evaluated were captured while the drone was hovering at a distance of 30 and 280 meters from the microphone array. We divided each signal into 100 segments of 200 ms, allowing the estimation of the delays between sensor pairs. We collected 6 candidate delays for each $r_{x_i x_j}$ forming a matrix $\mathbf{V}_{N \times 6}$.

In contrast to simulations, where we had theoretical delays among other delays,

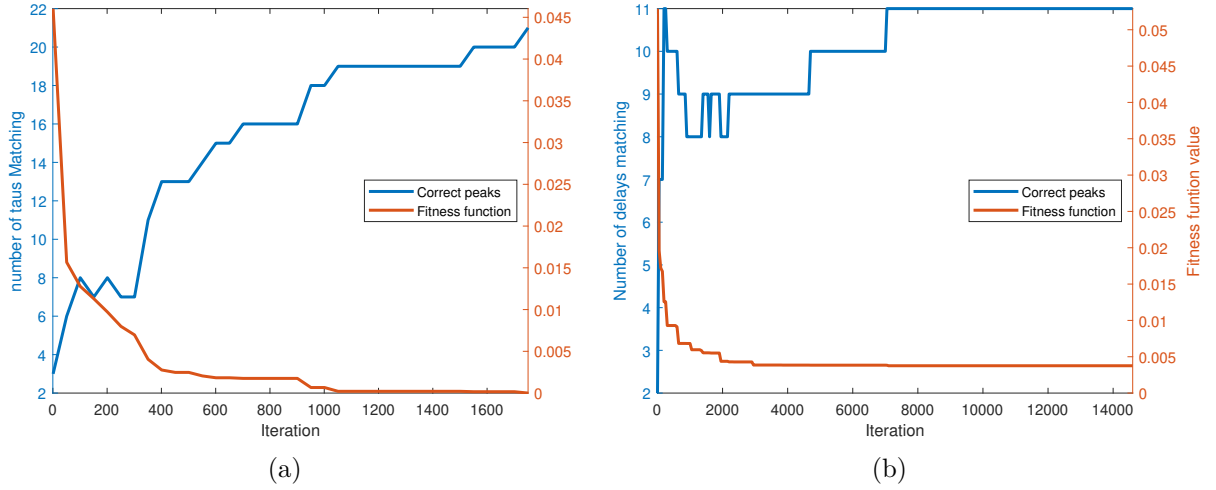


Figure 5 – GA search progress with fitness function and number of correct delays evolution. (a) 1600 iterations, and the search stopped according to the fitness function criteria (b) GA and the convergence to a local minimum.

it is highly unlikely for the peaks of an actual cross-correlation to correspond perfectly with the theoretical delay. Therefore, we monitored the progress of the GA by assessing the delay error (in samples), which is calculated as $\sum_{i=1}^N |\hat{\mathbf{v}}_i - \mathbf{v}_i|$, where $\hat{\mathbf{v}}$ represents the vector of estimated delays mapped by the GA, and \mathbf{v} consists of the theoretical delays.

Figure 6 illustrates the delay error for each analyzed window. The simulation was conducted using a total of 2,000 iterations, starting from a randomly initialized population. Specifically, Figure 6 (a) showcases the delay error progression for each 200 ms window, and Figure 6 (b) highlights the benefits of utilizing the GA algorithm in mitigating large delay errors.

The results depicted in Figure 6 denote the minimum delay error, i.e., the minimum possible error if we choose the element of matrix \mathbf{V} that minimizes the error of the TDE. The simulation results show that GA-ZCS has the potential to find all theoretical delays, so one challenge is to develop a method that finds an accurate delay for each pair of microphones.

Introducing a signal enhancement technique as a step in this method may enhance the signal of interest, specifically the drone signal, in each channel. By applying signal enhancement techniques, it is possible to improve the detectability and accuracy of the drone signal, thereby enhancing the performance of the time delay estimation method. Exploring such signal enhancement techniques holds promise for further improving the robustness and effectiveness of the overall approach.

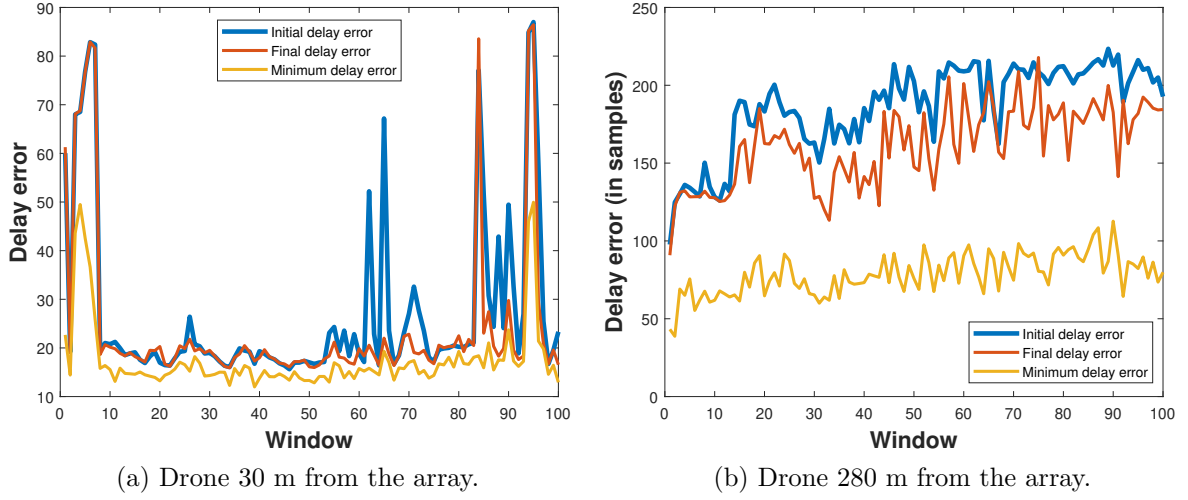


Figure 6 – Progress of GA search with fitness function, minimum delay error in samples, and the error reduction highlighting the benefits of the heuristic search. (a) Analysis of 100 windows using GA search progress while the drone hovers close to the microphone array (30 m); and (b) Analysis of 100 windows with the drone hovering at a distance of 280 m from the microphone array.

4.2 Exhaustive search with ZCS-LS

Here we present the results of acoustic-based drone DOA estimation using a 4-microphone array. This experiment with a reduced number of microphones allows a better comprehension of the effectiveness of the ZCS cost function using the ES approach.

To illustrate the versatility and robustness of our approach, we apply it to a challenging scenario utilizing a 4-microphone array setup. Through experimentation using both simulated and real-world data, our research underscores the potential of our novel DOA estimation methodology, showcasing its efficacy across diverse applications. We believe that these experiments prove the applicability of the ZCS-LS to near-real-time applications.

4.2.1 Data acquisition

In Figure 7, the Phantom 4 drone, an array of Behringer ECM8000 [136, 137] microphones, and the Zoom F8 recorder are illustrated, the latter serving to convert analog signals into digital format and store them in the wave format. The data acquisition process involves capturing acoustic signals emitted by the drone, along with the background noise. Following this real-time acquisition, the dataset undergoes a complete offline analysis to extract valuable insights and draw informed conclusions.



Figure 7 – Drone data acquisition set-up.

4.2.2 Drone noise collected with spherical array

Figure 8 offers a graphical depiction of both the background noise and the signals emitted by the hovering Phantom 4 drone, illustrating a duration of 500 ms. The figure also includes a spectrogram computed with a sample rate of 48 kHz. Notably, in optimal conditions and when the drone is close to the microphone array, it becomes feasible to capture drone noises in higher frequency ranges, extending up to 13.5 kHz. As the drone moves farther from the microphone, the noise becomes increasingly concentrated, primarily within the frequency range below 5 kHz.

4.2.3 Effects of signal window length

Before conducting simulations, a full evaluation of the actual signals was undertaken to achieve a more faithful emulation of real-world conditions. This experimental assessment is imperative to understand the frequency with which primary and secondary peaks accurately indicate the true time delay. Figure 9 illustrates the histogram detailing the position of the true time delay. Specifically, it shows the number of peaks sorted in amplitude descending order, highlighting instances where the correct time delay is successfully retrieved (with a permissible error of ± 1 sample). This scrutiny serves as an important step in ensuring the potential to explore the secondary peaks to estimate DOA.

4.2.4 DOA estimation with simulated data

To run simulations, we generated synthetic \mathbf{V} matrices according to τ_{max} between each pair of microphones. Figure 10 describes the cost function computed with simulated

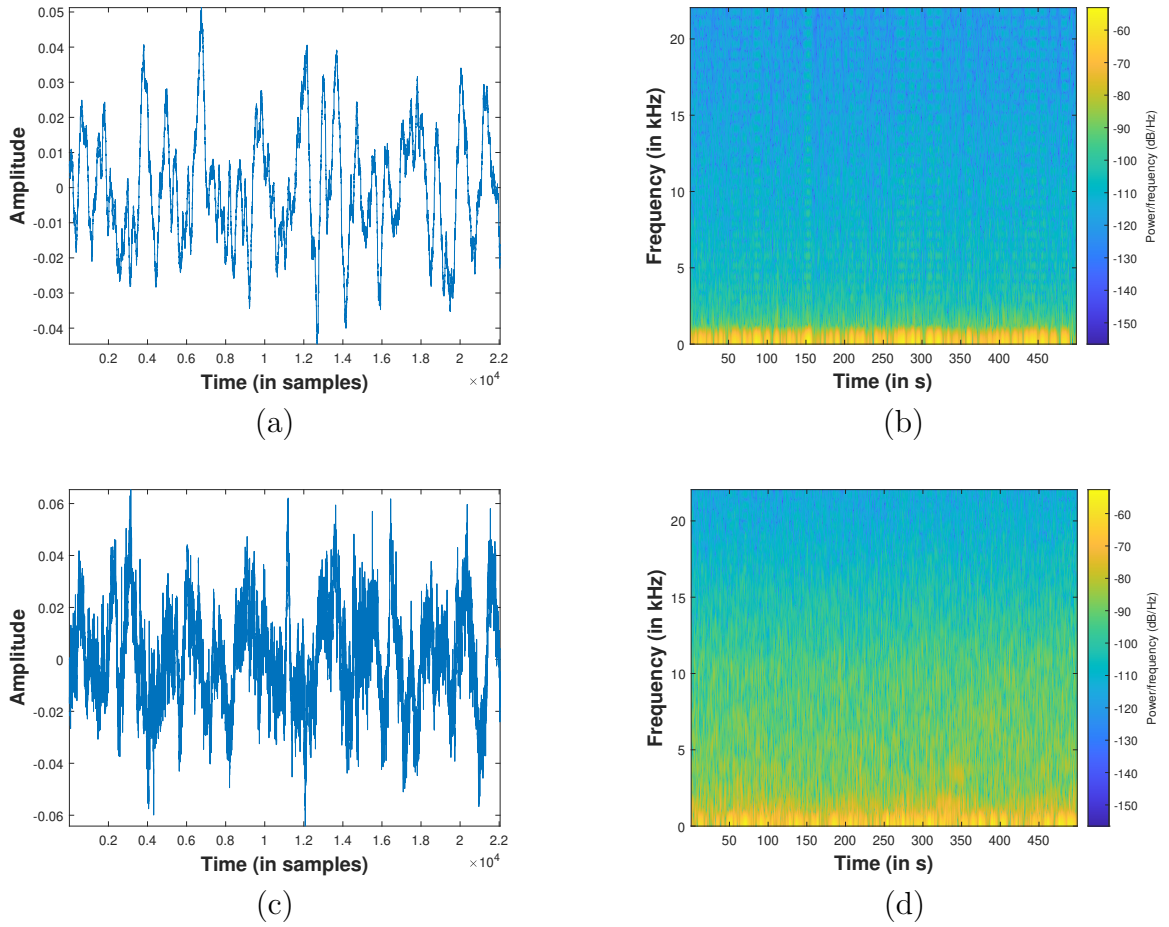


Figure 8 – Spectrogram of drone noise collected from different distances. (a) Background noise time domain (b) Spectrogram background noise (c) Drone noise time domain (30 m away from the microphone array) (d) Drone noise spectrogram (30 m away from the microphone array).

data.

The ZCS cost function facilitates the computation of all S potential combinations of time delays presented in the \mathbf{V} matrix. Figure 11 illustrates 1,000 runs with simulated data, displaying the position at which the correct time delay vector is situated according to the ZCS. Although the ZCS itself does not determine the optimal combination of time delays, Figure 11 illustrates that this cost function effectively places the correct time delay vector among the $Z = 100$ vectors, thereby reducing the solution space $S = C^N$ to 0.01%. With a streamlined solution space, we can calculate all Z vectors using the LS cost function, a more computationally complex method to further refine the estimation.

Figure 12 depicts the DOA estimation results for 1,000 trials. The accuracy obtained in this experiment with simulated data for the classic estimations (LS solution with no optimization technique) is 0%, ZCS-LS = 74%, and ZCS = 34%.

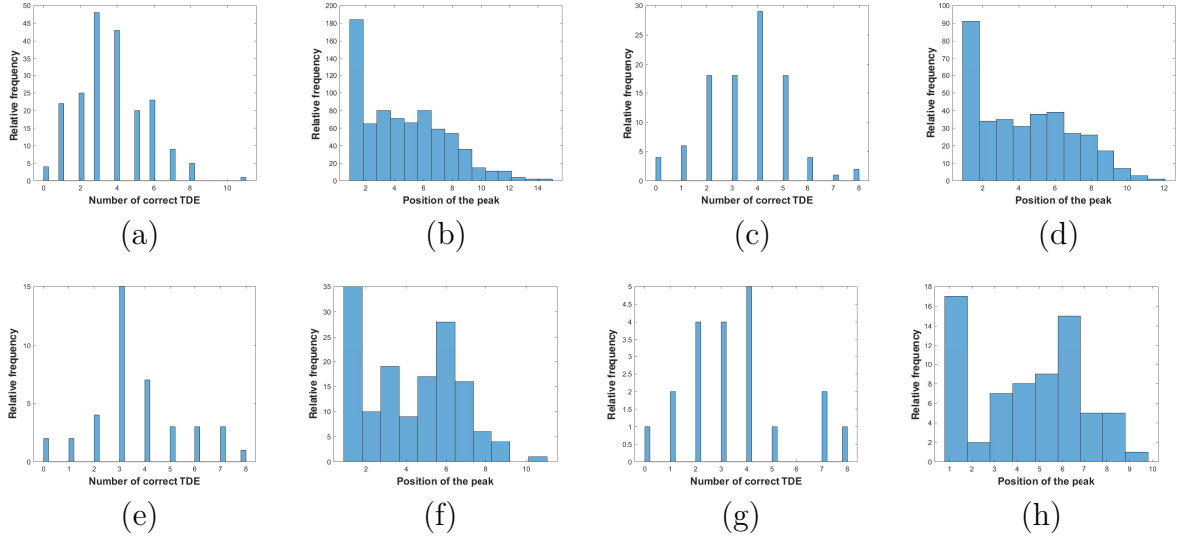


Figure 9 – The statistics of the cross-correlations $r_{x_i x_j}$. They reveal the number of accurately estimated delays (comprising all peaks from $r_{x_i x_j}$ and considering ± 1 sample error) within distinct time windows of 100 ms, 200 ms, 500 ms, and 1,000 ms for cases (a), (c), (e), and (g) respectively. In addition to the number of accurately estimated delays, the statistics of the cross-correlations $r_{x_i x_j}$ provide insights into the peak position based on the descent amplitude criterion within specific time windows of 100 ms, 200 ms, 500 ms, and 1,000 ms for cases (b), (d), (f), and (h) respectively.

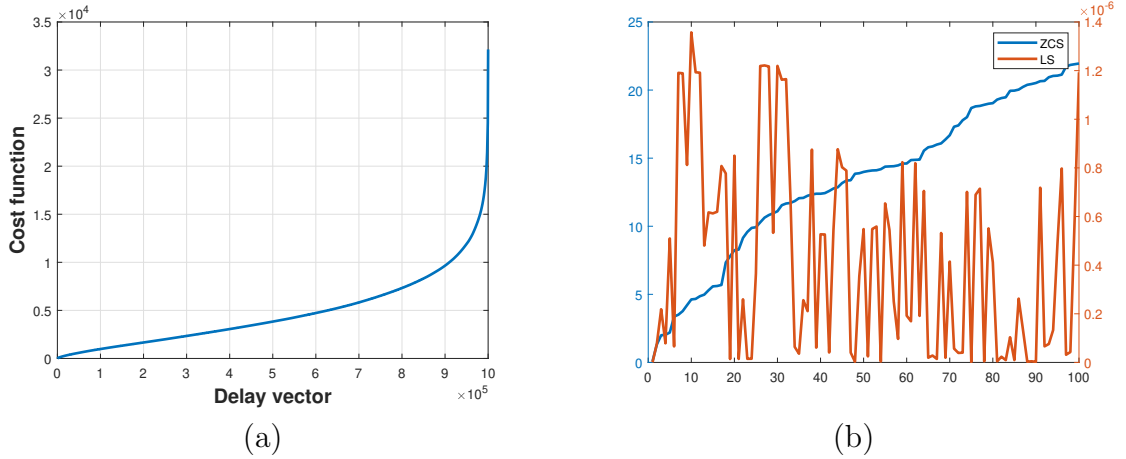


Figure 10 – Evolution ZCS and LS cost functions. (a) All delay combinations are calculated using ZCS and sorted in ascending order according to the ZCS cost function. (b) First $Z = 100$ ZCS and LS cost function.

4.2.5 DOA estimation with experimental data

The acoustic signals emitted by the drone often reach the microphone array accompanied by various distortions and challenges. A frequent error encountered in time delay estimation based on cross-correlation arises from the existence of noise. When the

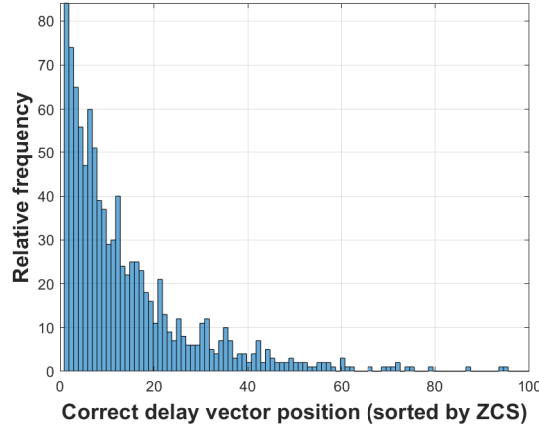
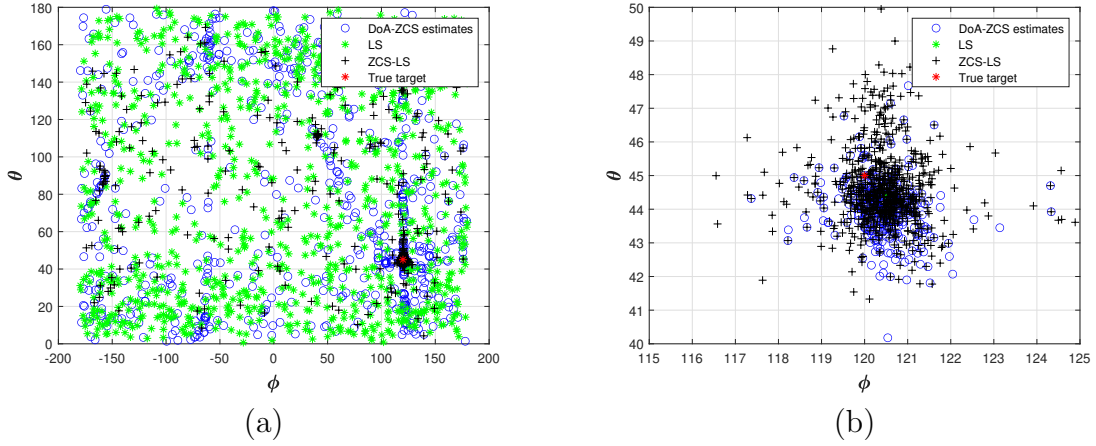


Figure 11 – Histogram of the position of the correct set of delays (1,000 trials).

Figure 12 – Comparison among LS, ZCS, and ZCS-LS DOA estimations (1,000 simulations). (a) results plotted for all possible θ and ϕ ; and (b) focused view within $\pm 5^\circ$ of the true target angles (θ, ϕ) .

correlated signals are affected by noise, it has the potential to introduce false correlations, causing inaccurate time delay estimations. This noise can distort the cross-correlation function, thereby causing misleading peak positions or the appearance of false peaks with high amplitudes that do not align with the actual time delay of the signal of interest.

Inaccurate time delay estimation can also originate from the existence of reverberation or multipath within the signals. The presence of reverberation markedly impacts the form and strength of the cross-correlation function, complicating the precise identification of the genuine peak denoting the direct path time delay. The reflections and diverse pathways of sound propagation may generate extra peaks or alter the primary peak, resulting in erroneous estimations.

Figure 13 illustrates pertinent issues associated with TDE when SNR is low. In Figure 13 (a), we observe the accurate TDE in the 5th peak, sorted by descending amplitude

order. Figure 13 (b) showcases a worst case in which only the 9th peak corresponds to the correct time delay. In conclusion, the cross-correlations encapsulate the requisite information for accurate DOA estimation; however, the efficacy of these estimations is compromised by the low SNR of the target signal.

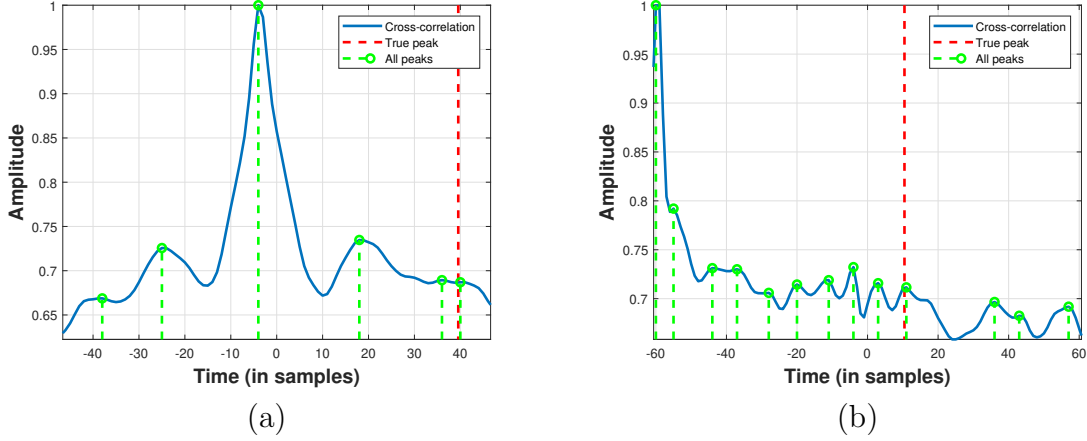


Figure 13 – Cross-correlations of the acoustic signals collected from a Phantom 4 drone hovering in an outdoor environment. (a) 5th peak corresponds to the correct time delay (b) 9th peak corresponds to the correct time delay

Figure 14 depicts the relationship between DOA error and TDE error, providing valuable insights into the accuracy of the localization process. The graph illustrates that as TDE estimation error increases, there is a corresponding rise in DOA error, indicating a direct correlation between the two parameters. Notably, the analysis reveals that TDE errors within a maximum range of three samples remain acceptable, as they correspond to DOA errors of less than 5 degrees for both zenith and azimuth angle estimations. This observation underscores the robustness of the localization system, suggesting that minor deviations in TDE estimation do not significantly compromise the accuracy of DOA predictions within a reasonable margin.

Figure 15 illustrates the drone DOA estimation results using different window sizes. The classic LS estimation approach (GCC-PHAT using only the primary peaks) yielded 0% accuracy, indicating poor performance in handling the complexities of the acoustic environment. In contrast, the ZCS and ZCS-LS methods pointed toward the correct direction, achieving an accuracy of $\text{ZCS} = 83.5\% \pm 3.6\%$ and $\text{ZCS-LS} = 90.2\% \pm 4.4\%$. Using ZCS and ZCS-LS results in a high-density area of estimations around the actual angles, $\theta = 10$ and $\phi = -25$. It should be noted that the ZCS facilitates exhaustive computations of all possible delay combinations, and the additional computation of the LS cost function enhances estimations by approximately 7%. The experimental outcomes surpassed the simulation ones, primarily attributable to the variable number of delay candidates (C) encountered. While the simulation phase maintained a fixed value of

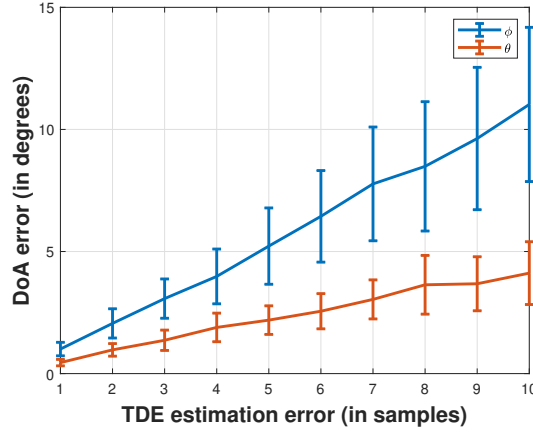


Figure 14 – Error of DOA according to the TDE additive noise.

$C = 10$, the experimental phase yielded a fluctuating range of peaks, ranging from 1 to 8 for each cross-correlation. This variability in delay candidates in the experimental setting contributed to the enhanced performance observed, demonstrating the method's adaptability in real-world scenarios.

The results presented in this Chapter demonstrate the effectiveness of the proposed approaches for DOA estimation in highly noisy and reverberant environments. Through both simulated and real-world experiments, it was shown that the ZCS cost function, particularly when combined with the LS cost function, significantly enhances the accuracy of delay estimations (even in scenarios where multiple peaks appear in the cross-correlation function). Additionally, both the exhaustive search and heuristic optimization using genetic algorithms proved capable of improving performance while managing computational complexity. These findings confirm the practicality of the proposed methods, validating their potential for real-time application in acoustic-based counter-drone systems.

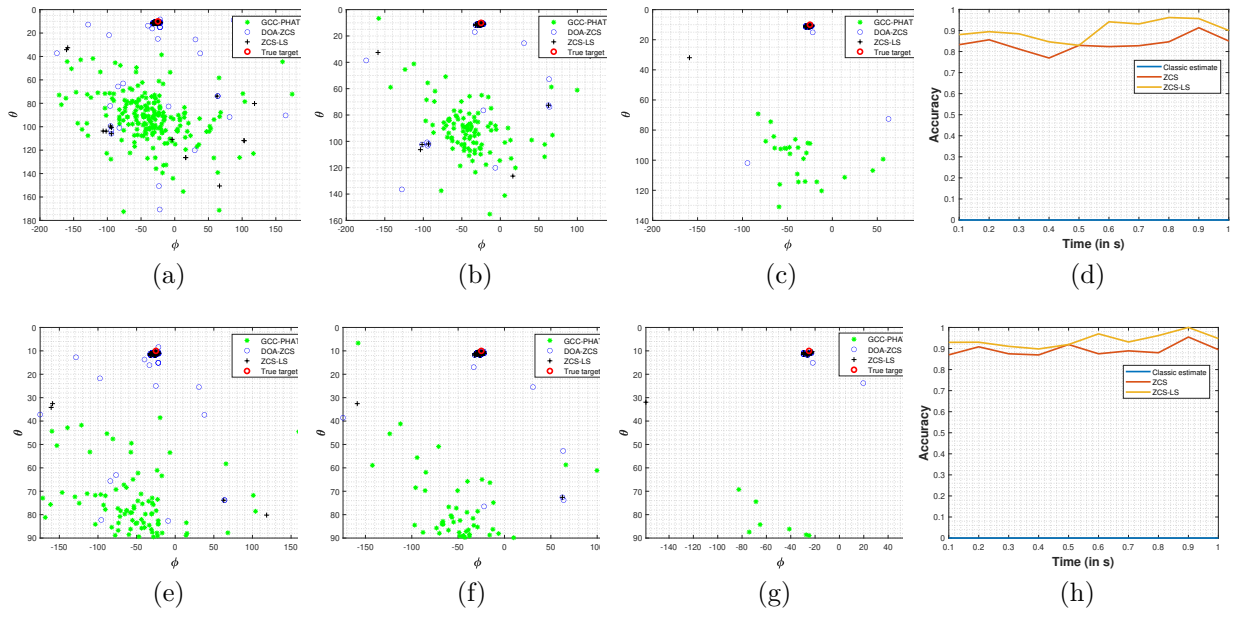


Figure 15 – Experimental results: drone DOA estimates 100 m away from the microphone. (a) 209 estimations with 100 ms windows (ZCS-LS accuracy of 88.0%) and (e) 100 ms windows (ZCS-LS accuracy of 92.9% discarding estimations pointing to the floor); (b) 104 estimations with 200 ms windows (ZCS-LS accuracy of 89.4%) and (f) 200 ms windows (ZCS-LS accuracy of 93.0% discarding estimations pointing to the floor); (c) 104 estimations with 800 ms windows (ZCS-LS accuracy of 96.1%) and (g) 800 ms windows (ZCS-LS accuracy of 96.1% discarding estimations pointing to the floor); (d) Accuracy of DOA estimators with different signal window sizes; and (h) Accuracy of DOA estimators with different signal window sizes, considering only estimated DOAs within the valid zenith angular range from 0° to 90° .

5 LOCALIZATION RESULTS

This chapter presents the results of drone localization estimation using both NN and TDOA-based approaches. The primary objective of this investigation is to evaluate the impact of acoustic reflections on localization accuracy. Section 5.1 analyzes the effects of reverberation on the cross-correlation functions. Section 5.2 reports the performance of NN- and TDOA-based localization techniques under reflective conditions. Finally, Section 5.3 discusses the key findings and insights derived from these experiments.

5.1 The Reverberation Effect

In acoustics, TDOA techniques use cross-correlation to estimate the delay between two signals, $x_i(k)$ and $x_j(k)$, from any pair of sensors (i, j) . Given M sensors, we can estimate the position based on C_2^M measurements [138, 60, 139], i.e., employ all $M(M-1)/2$ available pairs [61]. However, strong reverberation, usually present in indoor environment scenarios, causes wrong TDOA estimates. In such cases, the room impulse response (RIR) shows other components besides the line of sight (LOS) component, which may or may not be present depending on whether the environment presents obstacles to propagation.

Figure 16 illustrates the effect of reverberation in the RIR and the cross-correlation. In a non-reverberating room, where the clean signal is denoted as $s(k)$, the impulse response $h_1(k)$ shows only the LOS component, whereas, in a reverberating room, several other peaks are present in the RIR $h_2(k)$, with a peak being larger than the LOS component. Assuming a noiseless scenario, the signals from two microphones can be modeled as

$$\begin{aligned} x_1(k) &= s(k) * h_1(k), \text{ and} \\ x_2(k) &= s(k) * h_2(k), \end{aligned} \tag{5.1}$$

where “*” denotes the convolution operator.

The largest peak of the cross-correlation function $r_{12}(\tau) = E[x_1(k)x_2(k-\tau)]$ provides the TDOA, which can be estimated using any of the GCC algorithms [140], e.g., the GCC-PHAT. For instance, assuming additive white noise, an estimate of the conventional cross-correlation algorithm yields $r_{12}(\tau) = h_1(\tau) * h_2(-\tau)$.

As illustrated on the bottom left of Figure 16, GCC renders a clear peak with the correct TDOA of -2000 samples in the case of a non-reverberating room. Conversely, in the case of a strong reverberating room (right side of Figure 16), RIRs, $h_3(k)$ and $h_4(k)$ may exhibit other more prominence peaks than the LOS peak, thereby resulting in a cross-correlation $r_{34}(\tau)$ with multiple strong peaks, including eventually outliers. Therefore,

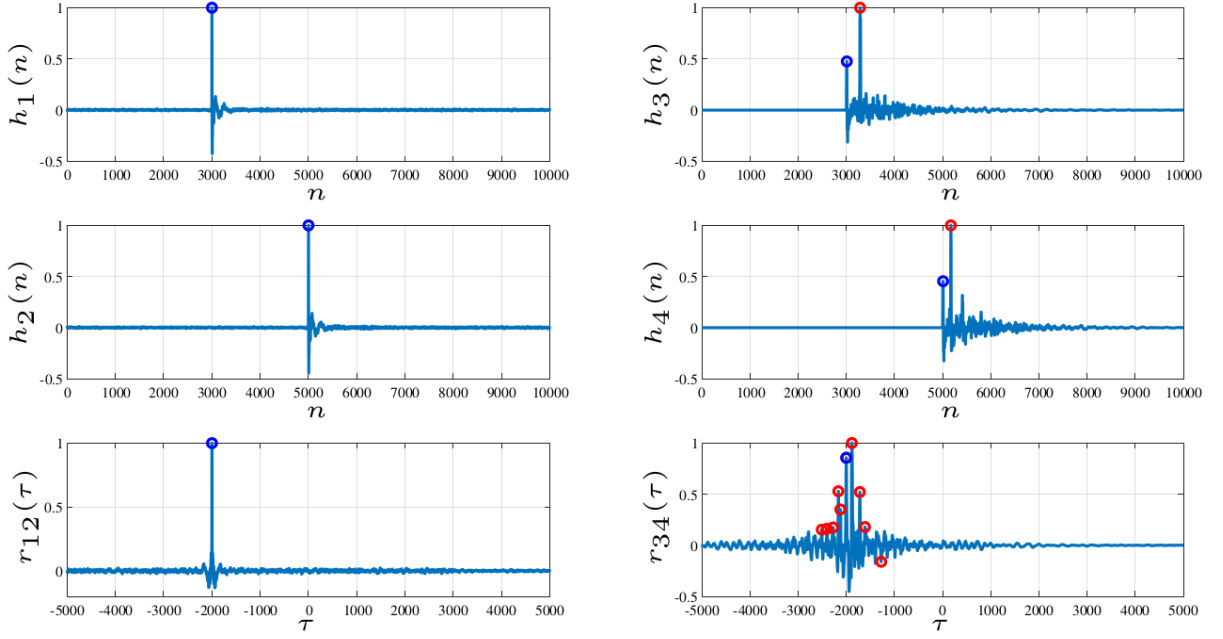


Figure 16 – Correct estimation of TDOA in a non-reverberating room (left) versus the effect of reverberation (right) on the RIR and the cross-correlation of the received signals. Source: [65].

in a practical indoor source localization problem, even a data-selection approach [141] would not be able to discriminate outliers, making the solution unreliable.

5.2 Localization Experiments

We conducted two experiments, each targeting distinct signals of interest (SoI) in different environments. The first experiment took place in a room using a speaker in a fixed position emitting a frequency-shift keying (FSK) modulated sequence, while the second one was carried out in a larger indoor space using the noise generated by a drone. The motivation behind these experiments is to first investigate the localization of a stationary acoustic emitter, establishing a performance baseline. Following this analysis, the focus shifts to estimating the position of a drone in motion (drone hovering with noisy positions due to its movement during recordings), allowing for the evaluation of localization methods under dynamic conditions. Detailed descriptions of these setups can be found in subsection 5.2.1 and subsection 5.2.2.

5.2.1 Fixed synthetic acoustic emitter experiment

We performed several experiments using an audio signal comprising two seconds of white Gaussian noise, followed by a 60-second-long SoI, a 100 bps binary FSK modulated sequence centered at 15 kHz. However, the experiment described here focuses only on the

SoI. Figure 17 shows time and frequency-domain representations of the complete audio signal, featuring a magnified segment of the SoI (top right) transmitted from distinct speaker positions and captured by eight microphones close to the room walls.

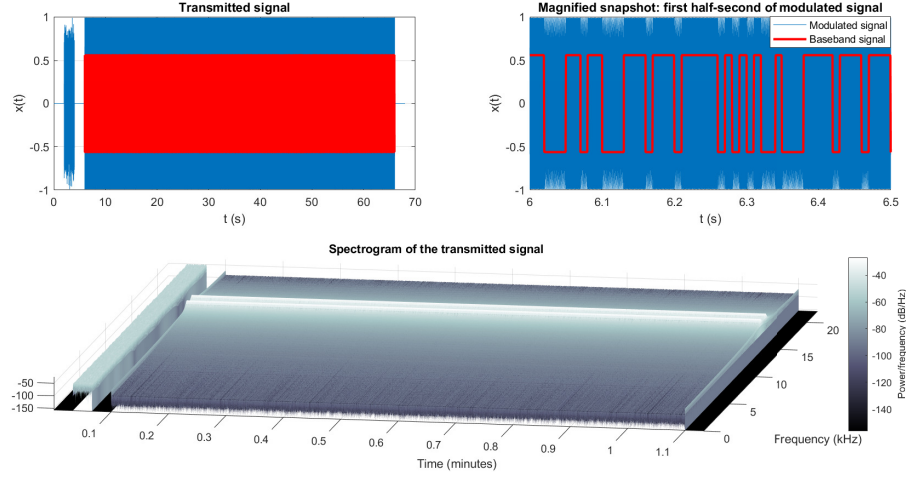


Figure 17 – Spectrogram of the audio signal employed in the practical experiments.
Source: [65]

We used a substantially reverberating classroom, with xyz -dimensions of $8.32\text{m} \times 5.29\text{m} \times 3.94\text{m}$ for the recordings; assuming room walls with an absorption coefficient of 0.1, we estimated a reverberation time T_{60} , the time it takes for the sound level to drop 60 dB after sound cessation [142], as 1.43. Figure 18 illustrates the positions of the target speakers and the microphones. The recorded signals also contain an ambient noise component; however, given the inherent ability of the proposed method to cope with background noise, we shall focus our attention on the reverberation effect, which is more critical in introducing outliers.

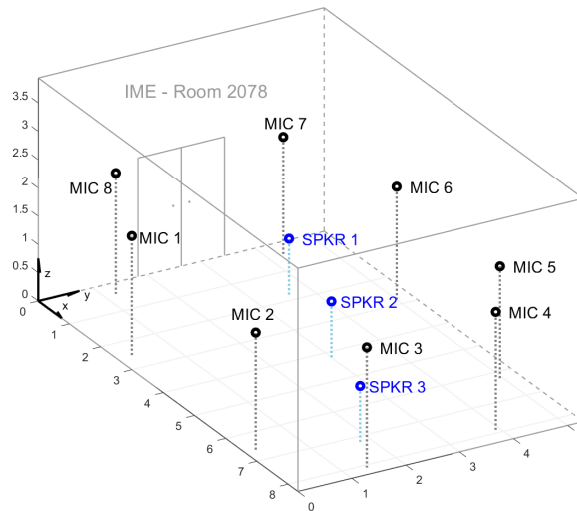


Figure 18 – Classroom with eight microphones (MIC 1 to MIC 8) and three distinct speakers (SPKR 1, SPCR 2, and SPCR 3).

We tried different hold-out setups when training the NN to achieve good results. Table 9 shows the results of the first experiment. Here, the mean squared error (MSE) is calculated based on the difference between the estimated and the actual position, varying the block time and resolution parameter (RP). We consider the results excellent, given the strong reverberation in the room, especially for window sizes of 500 ms or larger. Comparatively, classical TDOA-based methods, known to perform poorly under such conditions, were unable to achieve comparable levels of accuracy.

Table 9 – Neural network regression results

Block (time)	MSE			RP η
	Training	Validation	Testing	
1 s	$1.95e-15$	$2.64e-9$	$4.98e-10$	100
500 ms	$2.46e-15$	$3.49e-9$	$2.16e-8$	50
250 ms	$2.61e-15$	$1.49e-2$	$1.68e-4$	5

The 1-minute size of the transmitted signal, in the case of 1-second blocks, leads to 180 observations, 60 for each of the three positions, of $K = 560$ features (ten peaks for each $N=28$ TDOAs and respective amplitudes). We estimated 28 TDOAs for each observation, a total of 5,040, from which we found 4,409 outliers. In other words, only 631 estimates were considered correct within ± 50 samples due to possible sensor and emitter positioning errors. That is equivalent to a maximum range difference error of 40 cm. The reverberation effect is so strong that, only in 23.3% of the cases (1,176 out of 5,040), the correct TDOA is within the ten most prominent peaks. For the case of 500 ms blocks, owing to the smaller block size, we had a dataset of 360 observations with $K = 560$ features. In this case, the number of outliers was 8,994 out of 10,080 estimated TDOAs, with 1,086 correct estimates, out of which 20.9% of correct TDOAs were within the dominant peaks. Finally, for the case of 250 ms blocks, a dataset of 720 observations, we found 19,451 outliers out of 20,160 TDOA estimates, with 709 correct estimates and 8.6% of correct TDOA estimates within the ten dominant peaks. We refrain from displaying the performances of the classical, extended, and data-selective LS algorithms, for they are unreliable, even in the case of the largest block of 1 s. Further details on these limitations can be found in the work of Apolinário Jr. et al. [141].

From this experiment, we realized that a network trained with a given block size behaves better when testing features estimated using the same block size. Furthermore, although a network trained with 1-second blocks performed well when testing features from 500-ms blocks, the converse did not hold.

We could not use the setup from the first experiment to generalize a regression and needed to train more positions. Hence, we extended the dataset to a rectangular grid of 104 positions, covering the classroom with loudspeakers equally separated by 60 cm, thereby increasing the number of classes. However, the performance of a NN

degrades with an increasing number of classes; consequently, the fitting problem for 500 ms blocks, even with a few adjustments, e.g., increasing the number of hidden neurons to 10, was disappointing. With a smaller dataset size, tested with ten positions, the Bayes regularization algorithm worked well, but was too slow for a 104-point grid. Changing the application from regression to classification did not favor the performance of the localization scheme as much as expected. The NN-based classification usually yielded better results, but even a few outliers would cause significant positioning errors.

After many approaches focusing on the NN, we decided to change the features for a more practical setup with a 104-point grid. We finally achieved better results after realizing that the ten most prominent peaks in the audio application failed to act as a fingerprint. In the event of bad results, we observed that the peaks were all clustered together in the principal elevation of the cross-correlation, failing to adequately describe the whole sequence due to the absence of samples from secondary hills. Nonetheless, we used the peak of the main elevations of the cross-correlation as the input for the neural network.

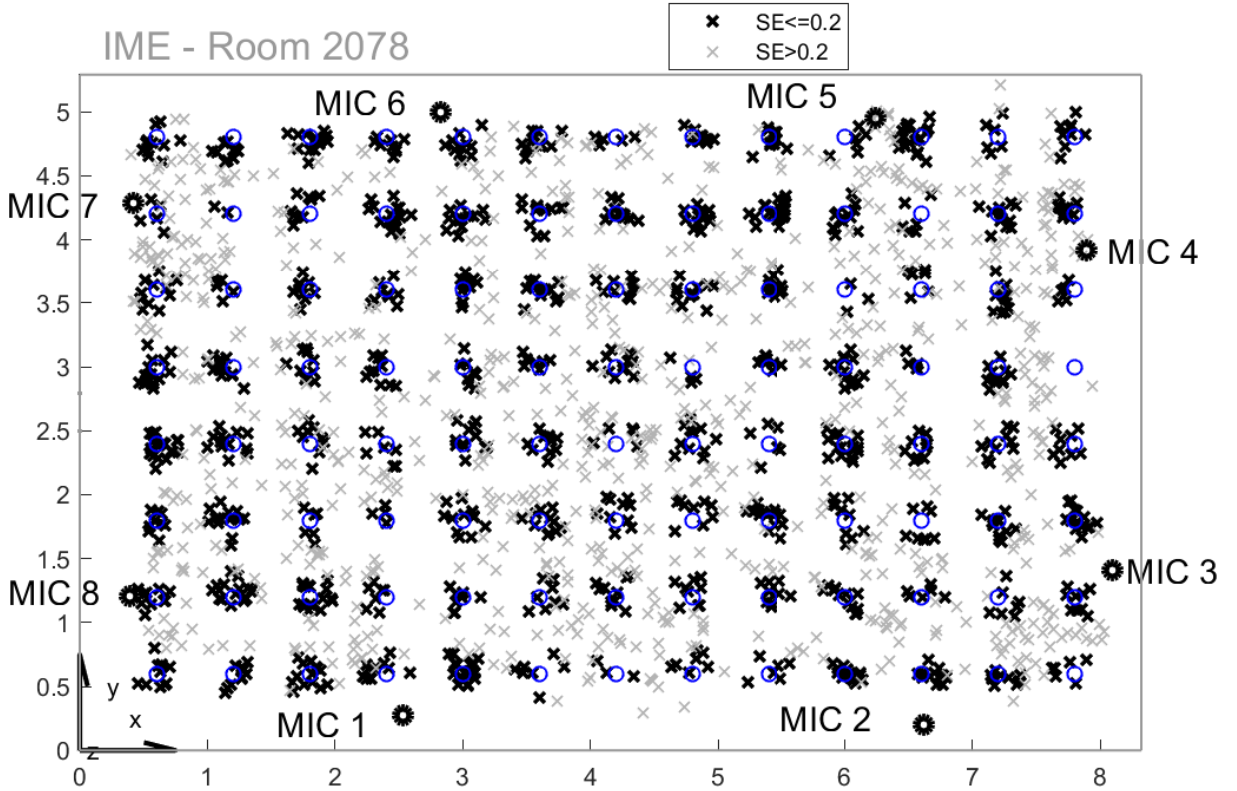


Figure 19 – Results for the position estimation.

To improve the results even further, we decided to change the structure of the NN of Figure 2. Thus, we exhaustively searched for the hyperparameters that minimize the MSE for this more complex problem. The updated NN is a Multi-Layer Perceptron (MLP) regressor, and the parameters that contributed the most to enhancing the performance are

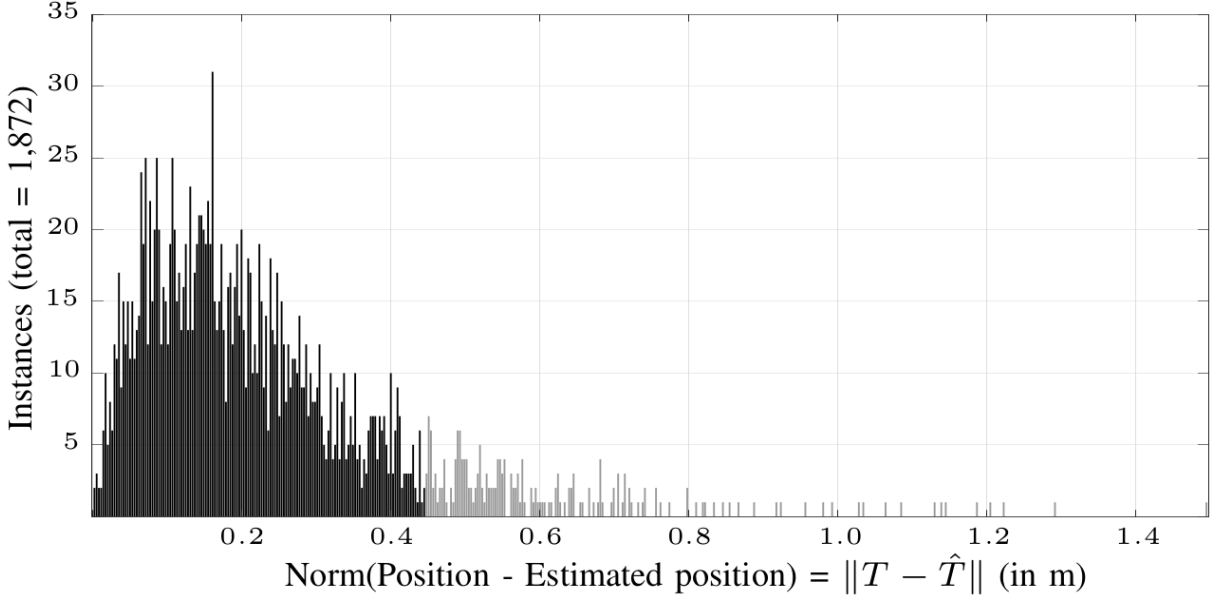


Figure 20 – Histogram of the estimation errors.

L (number of neurons) and the number of hidden layers. The best configuration achieved is a four-layer NN with three activation layers with an increased number of neurons $L = \{900, 600, 300\}$, and one output layer. Figure 19 and Figure 20 depict the result of the estimated positions and the histogram of the estimation errors, respectively, for all 104 positions in the test dataset. Note that the majority of the estimations (53.5%) presented errors below 0.4472 m (black bins of the histogram), i.e., the squared error (SE) below 0.2.

5.2.2 Drone localization Experiment

As noted in the previous experiment, the small room exhibited significant reverberation, resulting in the cross-correlation function showing numerous false peaks, mainly due to reverberation effects rather than the line-of-sight acoustic component. To address this, the trials described in the following were conducted in a larger indoor environment with walls and ceiling much farther from the microphones, but with a similar geometry of the microphones distributed along the terrain. This setup is intended to yield cross-correlations with primary and/or secondary peaks that correspond to the main acoustic component of the SoI (drone noise). Figure 21 describes the new setup used for drone noise recordings, which employs eight microphones.

The dataset created using this setup comprises five acoustic signals, each recorded using eight Behringer 800 microphones positioned around the SoI emitter in five different positions. In this case, the SoI is the acoustic signal generated by the propellers of a DJI Phantom 4 drone, captured at a sampling frequency of $f_s = 44.1$ kHz. Figure 22 depicts time-frequency representations of the drone noise caused by propellers and the background

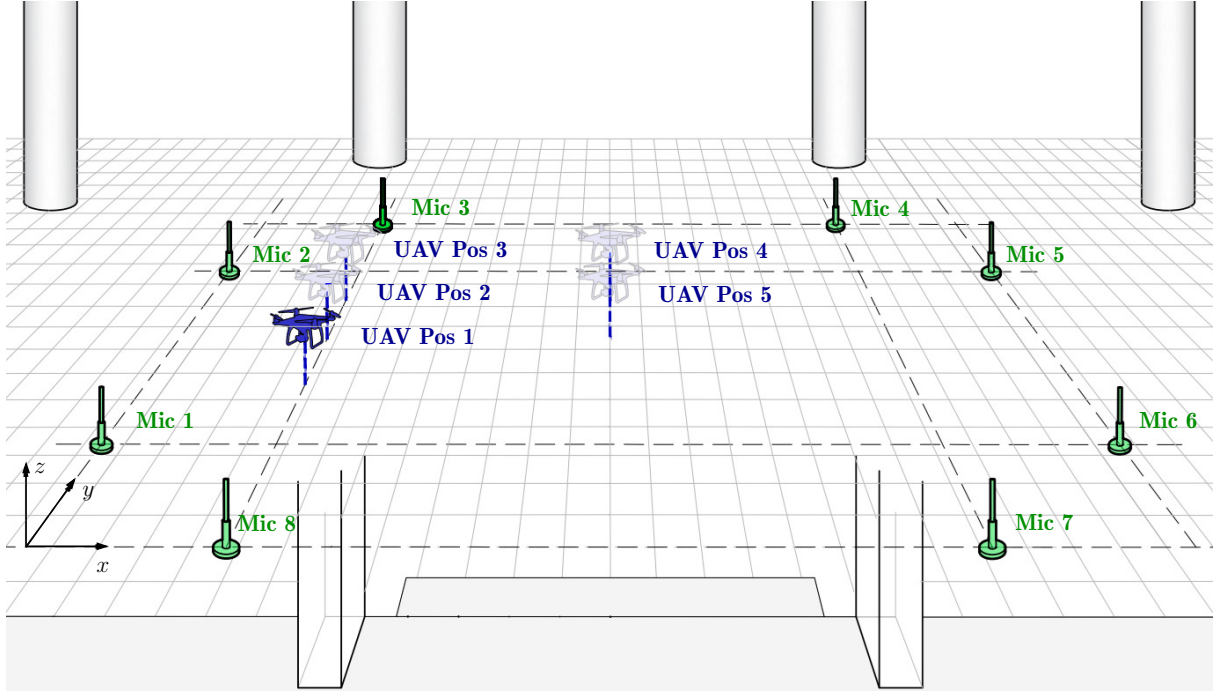


Figure 21 – Set-up for the drone noise recording.

noise recorded in this setup. It is possible to note that the drone noise is a broadband signal, covering the frequency region ranging from 20 Hz to 22,050 Hz.

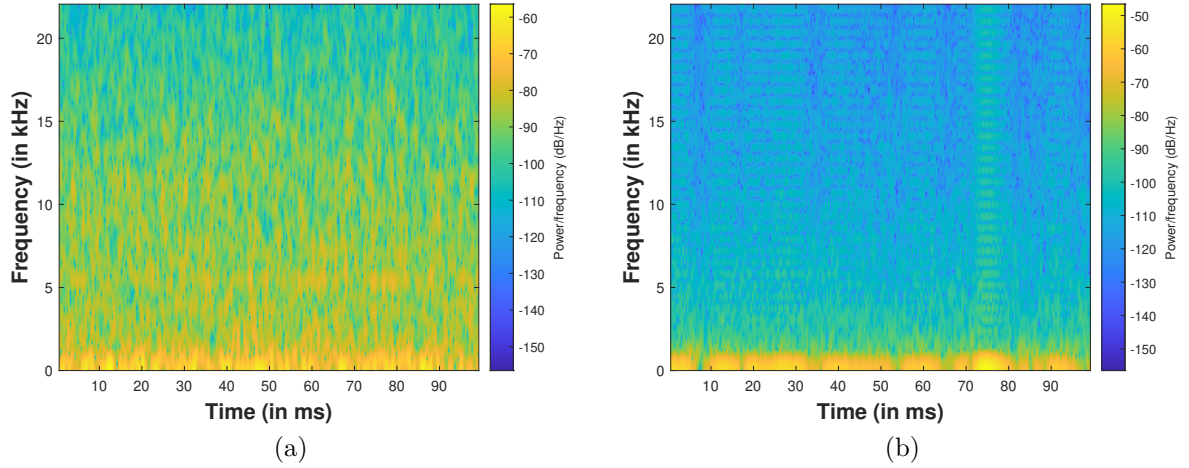


Figure 22 – Phantom IV drone spectrogram. **(a)** spectrogram with drone noise **(b)** background noise, i.e., all noises surrounding microphones except the drone noise.

Figure 23 depicts the geometry of the TDOA-based localization estimation methods. Δd_{12} is an example of one measure between pairs of microphones that can be explored to localize acoustic sources. This distance is estimated using the cross-correlation function [134].

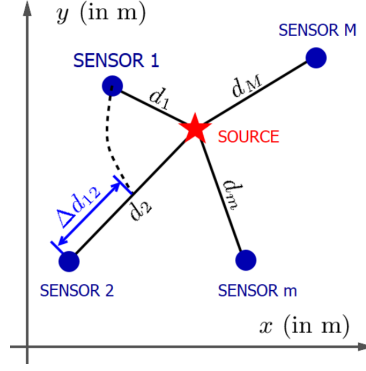


Figure 23 – Geometry of TDOA-based localization methods. Source: [141]

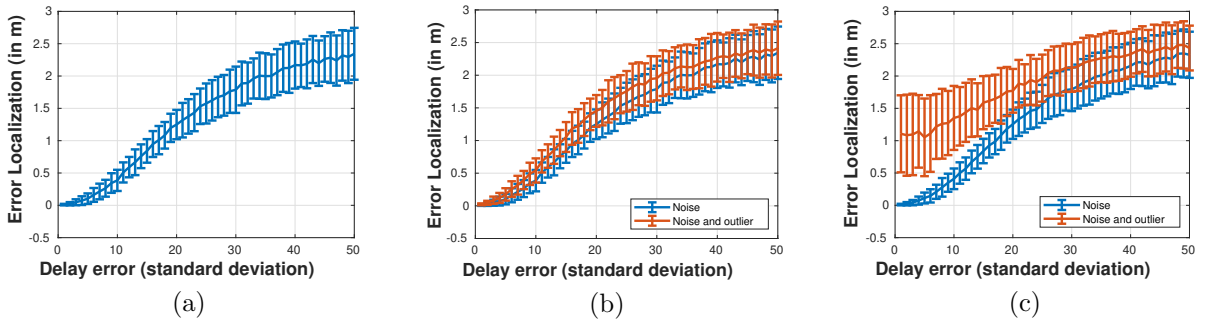


Figure 24 – Results obtained using simulated delays (in number of samples) with additive noise, based on the dataset’s geometry and assuming the drone is hovering at position five. **(a)** Localization error as a function of the standard deviation of the additive noise. **(b)** Localization error when introducing a single outlier, while varying the noise standard deviation. **(c)** Localization error when introducing two outliers, while varying the noise standard deviation.

Figure 24 depicts the behavior of localization error under TDOAs with additive noise and introducing one and two outliers using all TDOAs. The simulated TDOA values were derived from the actual microphone geometry and assuming the drone is hovering at position five. Figure 24 (a) shows the effect as the additive Gaussian noise increases, the localization error grows gradually, reflecting a predictable degradation in estimation accuracy due to uncertainty in the TDOAs. Figure 24 (b) denotes that introducing one outlier slightly increases the estimation error, i.e., this demonstrates that this system of equations has sufficient redundancy, and this one outlier is diluted among the many correct delays. Figure 24 (c) depicts the error rising sharply with low additive TDOA error due to the presence of two outliers. These results highlight the critical need for TDOA selection mechanisms that are not only noise-aware but also resilient to outliers, as few outliers can have a disproportionately large effect on the localization accuracy. The mean TDOA estimation error over 10 samples is approximately 0.5 m; however, the presence of two outliers increases the average error to 1.4 m, as illustrated in Figure 24 (c).

Overall, the results underline two key insights. First, while additive noise alone

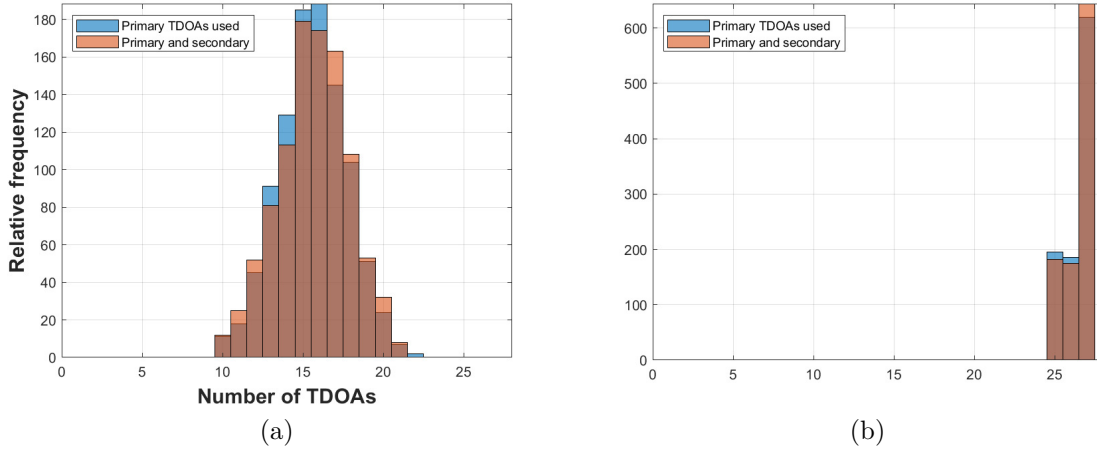


Figure 25 – Number of peaks used. (a) one pass GTS approach, and (b) GTS.

gradually degrades localization accuracy, it is manageable up to a certain threshold. Second, even a small number of outliers can severely increase the localization error. These findings reinforce the importance of incorporating strategies like the proposed ZCS-based cost function and delay selection algorithms to mitigate the influence of erroneous TDOA values and ensure more reliable localization performance under real-world conditions.

Figure 25 (a) depicts the relative frequency of TDOAs used to estimate the localization of the drone using the GTS technique. It should be noted that 15 is the largest number of TDOAs used. The minimum number of TDOAs is 10 up to 21, which means that the TDOA selection mechanism improves the results, but there is a minimum number of TDOAs that minimize the localization error. Figure 25 (b) depicts the relative frequency of TDOAs when GTS is used, iterating all TDOAs and choosing the one that has the highest contribution.

Figure 26 illustrates the cross-correlation issues that can arise in multipath environments. The cross-correlations shown pertain to the seventh and eighth microphones, with a maximum delay of 649 samples around the center (zero delay). Analyzing these cross-correlations can help researchers create new strategies to identify the most relevant peaks among numerous possibilities. Initially, a peak can be defined as a sample with a higher amplitude than its adjacent samples. However, this method may not effectively capture samples that span significant portions of the cross-correlations, potentially resulting in the selection of many too-close peaks. Thus, another processing is needed, i.e., identify the peak with a higher amplitude than its adjacent peaks.

Initially, we ran simulations to grasp the possible behavior of the localization estimation techniques with noisy TDOA, i.e., TDOA with additional error. Figure 27 shows the results of this first simulated experiment. Delay error is the mean number of samples added to the theoretical delay for each cross-correlation. These simulations

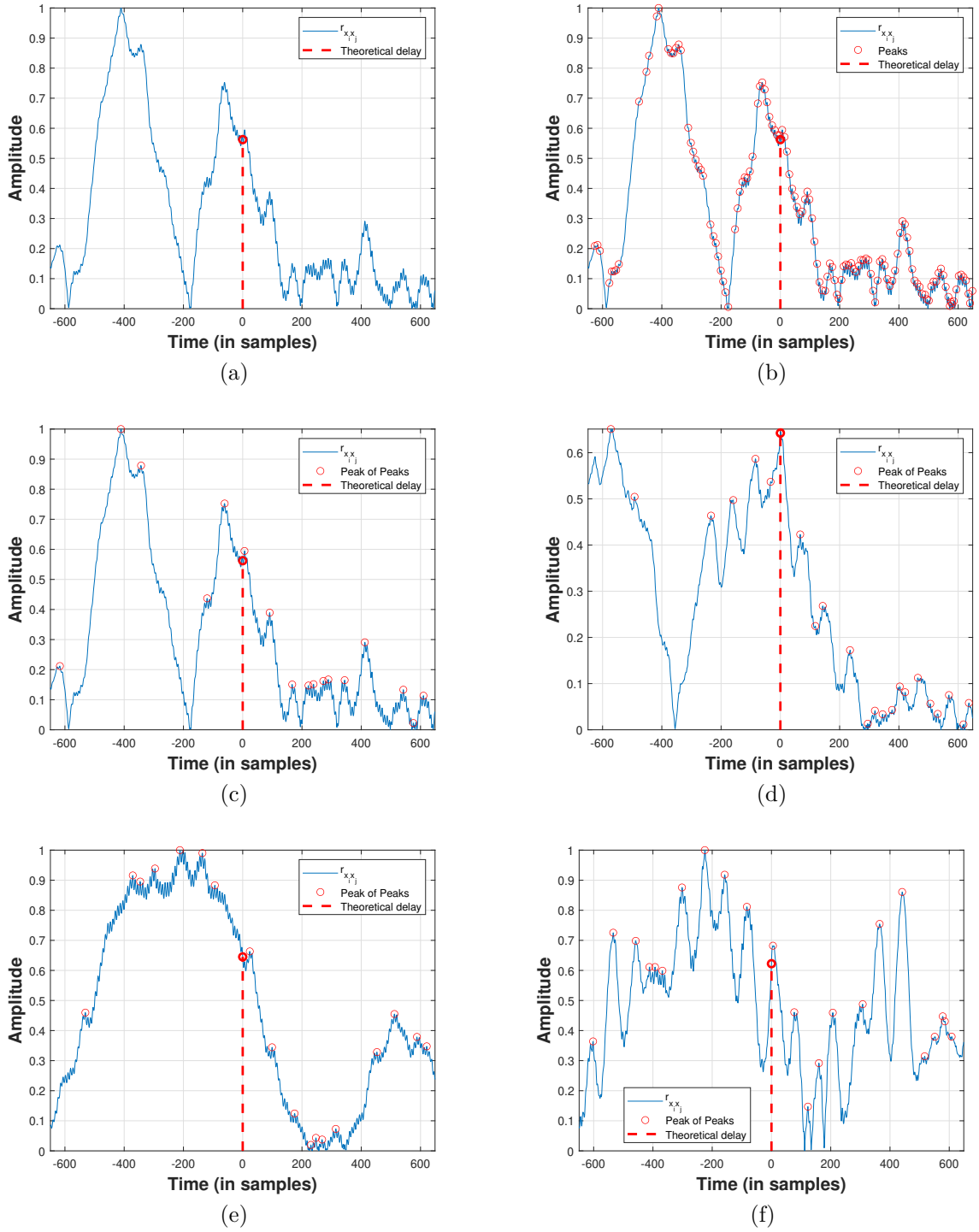


Figure 26 – Peak extraction process. (a) cross-correlation from microphones 7 and 8 with the drone hovering in position 5, (b) the same cross-correlation with the peaks detected, i.e., a peak is the sample that has a higher amplitude than the two nearest neighbors, (c) simple peak extraction method (sample amplitude greater than the other two closest peaks), the true peak estimation is the 4th highest peak, (d) theoretical delay is the peak with the highest amplitude, (e) theoretical delay is closer to the 7th peak, and (f) theoretical delay is closer to the 9th peak.

were designed to analyze the performance of localization techniques when the TDOA measurements are corrupted by noise, introducing additional errors into the system. By systematically adding varying noise levels to the TDOA values, we could observe how these errors propagate through both conventional and extended LS localization methods. This experiment allowed us to evaluate the sensitivity of each approach to TDOA inaccuracies and provided insights into the robustness of both techniques, particularly in environments where precise TDOA estimation is challenging.

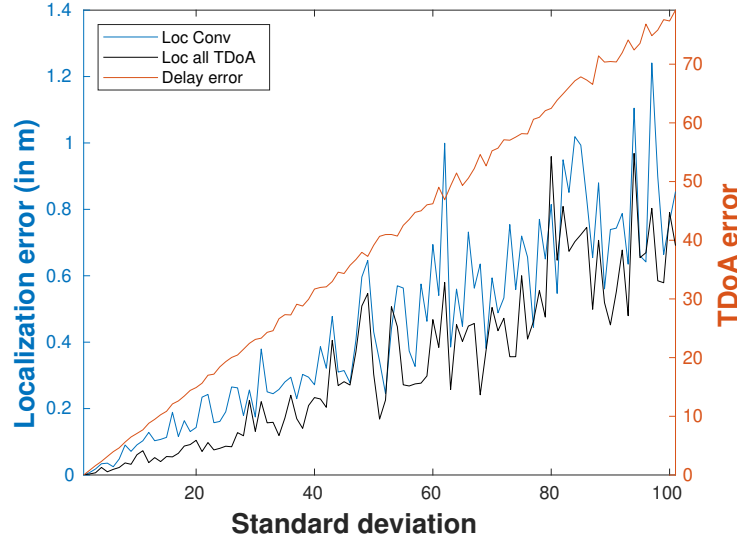


Figure 27 – Comparison of TDOA-based LS solutions using simulations.

Figure 28 (a) and Figure 28 (b) depict results obtained by using the main peak of the cross-correlation function between pairs of microphones. The delay error follows a Gaussian distribution with $\mu = 0$ and standard deviation σ , ranging from 0 to 40 (samples). The best results were obtained using the LS solution, considering the TDOA within the distance between each pair of microphones. The five 20-second recordings lead to 200 observations for each signal with a window of 100 ms. We estimated 28 TDOAs for each observation, a total of 5,600.

Figure 29 (a) and (b) depict results obtained by using the peak of the entire cross-correlation function between pairs of microphones (no restriction related to the known distance between microphones). However, this approach leads to localization estimates with significant errors, primarily due to inaccuracies in TDOA estimation. These errors become more likely when pairs of microphones capture uncorrelated signals, which tends to occur as the distance between microphones increases. The best results were obtained using the extended LS solution.

In this simulation, Figure 29 (c) and (d). the peak of the cross-correlation is restricted to the interval $(-\tau_{max}, +\tau_{max})$. This approach recognizes that the main peak may sometimes correspond to uncorrelated signals. By limiting the interval, TDOA errors are minimized, as correlated signals only produce peaks within the maximum possible time

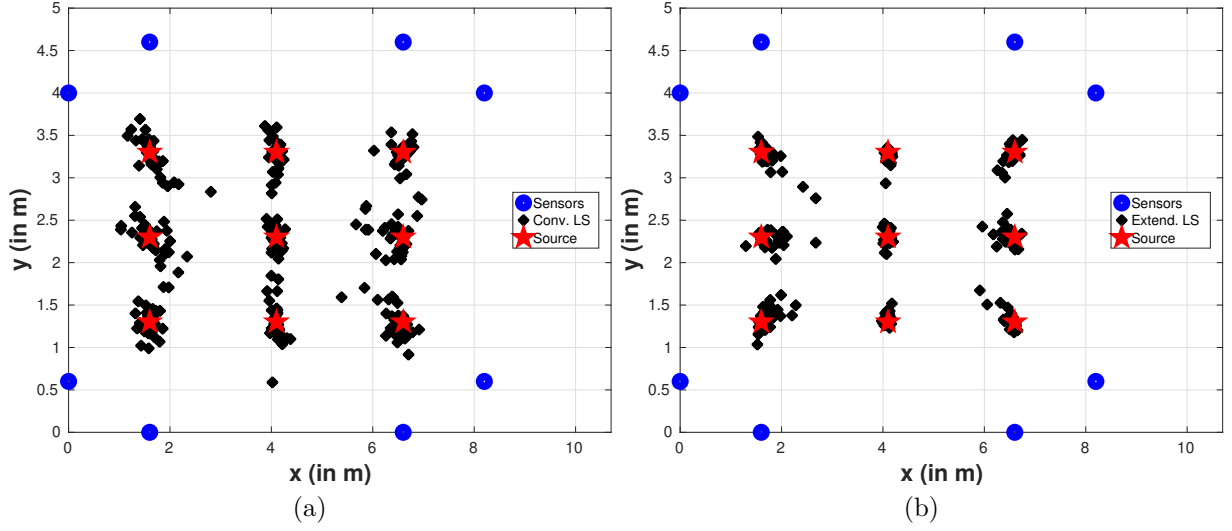


Figure 28 – Simulations for the drone localization problem. (a) conventional LS [60, 138, 139] (b) Extended LS [61].

delay. Any correlated acoustic signal between microphone pairs should generate a peak corresponding to the maximum Euclidean distance between the microphones. Again, the best results were obtained using the Extended LS solution [61] and the TDOA considering the distance between each pair of microphones.

5.2.3 Comparison of the localization techniques

In this section, we provide a comparative analysis of the TDOA-based LS solution and the NN approach for sound source localization. It is noteworthy that in the experiment with the synthetic acoustic emitter (Subsection 5.2.1), the results from the TDOA-based LS solutions were insufficient to establish a meaningful comparison. However, the experiment using drone noise in a larger environment (Subsection 5.2.2) enabled the TDOA-based approach to produce reliable results, making it possible to compare the two techniques. With this, we aim to highlight the strengths and limitations of each method in terms of accuracy and practical applicability for the drone experiment. The LS method, which estimates the source position based on TDOA measures between microphones, is straightforward and does not require a training phase. In contrast, the NN approach uses features from the environment’s reverberation fingerprint, allowing it to adapt and perform better in challenging scenarios with complex reverberation patterns. By evaluating both methods under the same conditions, we provided a better understanding of their respective performance and the trade-offs involved in choosing one technique over the other.

Table 10 presents the localization errors for the drone position estimation using three different methods: the conventional LS solution, the extended LS solution, LS-ES, LS-GTS, ZCS-LS, ZCS-LS-ES, ZCS-LS-GTS, and the NN approach. The results are reported

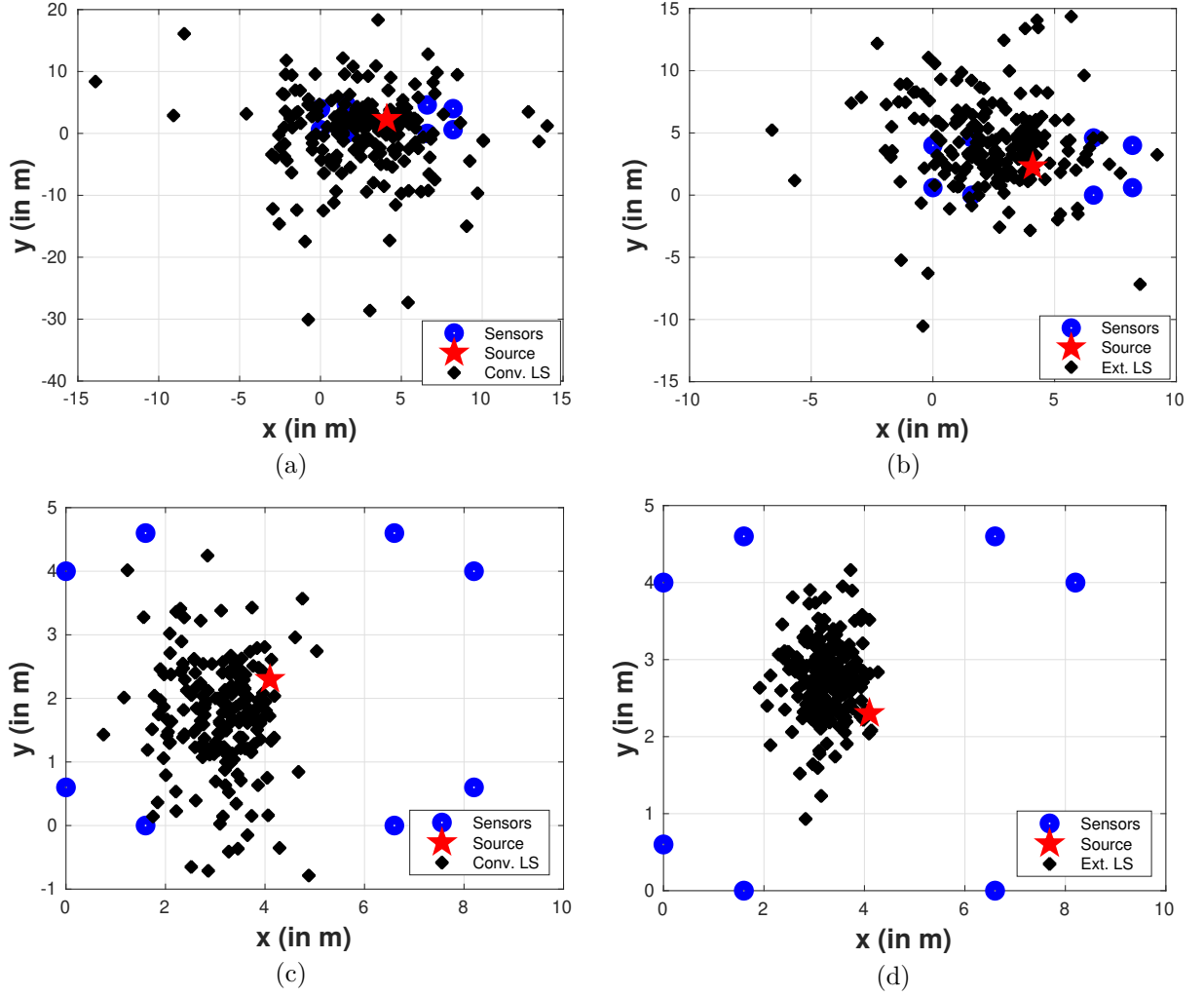


Figure 29 – Drone localization estimates. (a) conventional LS solution without taking into consideration τ_{max} (b) extended LS solution without taking into consideration τ_{max} (c) conventional LS solution with TDOA within τ_{max} (d) extended LS solution with TDOA within τ_{max} .

for the five different drone positions, with the error expressed as the mean localization error and the standard deviation. The best results are highlighted.

The conventional LS method exhibits relatively high localization errors, ranging from 2.25 to 3.43 units, with position 4 showing the largest error of 3.43 ± 1.10 . The extended LS solution shows a slight improvement, with errors generally lower than the conventional LS method, especially at position 4, where the error drops to 1.89 ± 0.60 meters. The most significant reduction in localization error is achieved using the NN approach. This method yields the smallest errors across all positions, with values as low as 1.02 ± 0.02 for position 2. This highlights the superior accuracy and stability of the NN method compared to both LS-based solutions. However, this superior accuracy comes at a cost, i.e., the NN approach requires a dataset for the training phase, which adds complexity in real-world applications where such data may not always be available. It

is also important to note that in the drone experiment, the drone lacked GPS, causing its position to vary slightly during recordings. This variability partially explains the less accurate results compared to the first experiment, where a speaker was carefully placed in fixed positions and remained stationary throughout the recordings.

Table 10 – Drone localization error and standard deviation

Drone Position	1	2	3	4	5
Conventional LS	2.36 ± 0.64	2.25 ± 0.84	2.78 ± 0.94	3.43 ± 1.10	2.86 ± 0.87
Extended LS	2.22 ± 0.35	2.40 ± 0.67	3.30 ± 0.75	1.89 ± 0.60	1.98 ± 0.25
LS-ES	2.04 ± 0.32	2.22 ± 0.37	2.83 ± 0.37	0.68 ± 0.39	1.21 ± 0.33
LS-GTS	1.87 ± 0.37	2.08 ± 0.45	2.73 ± 0.43	0.84 ± 0.40	1.38 ± 0.37
ZCS-LS	2.61 ± 0.38	3.05 ± 0.57	3.97 ± 0.61	1.17 ± 0.37	1.92 ± 0.15
ZCS-LS-ES	2.11 ± 0.35	2.27 ± 0.37	2.85 ± 0.38	0.55 ± 0.35	1.07 ± 0.36
ZCS-LS-GTS	2.00 ± 0.39	2.20 ± 0.43	2.79 ± 0.44	0.66 ± 0.40	1.21 ± 0.41
Neural Network	1.28 ± 0.01	1.02 ± 0.02	1.56 ± 0.02	1.69 ± 0.02	1.50 ± 0.02

The results shown in Figure 30 highlight how the number of selected TDOAs affects the accuracy and consistency of source localization. Each figure has 28 estimates that are the estimates with the lowest LS cost values. Figure 30 (a) and Figure 30 (b) show the results using only 5 and 6 delays (ES(5) and ES(6)). These configurations lead to low precision, with multiple scattered estimates due to the large number of possible combinations of delays that are mapped to a large number of different solutions. In contrast, Figure 30 (c) with 25 delays (ES(25)) shows a significant improvement, producing three dense and close clusters and reducing the spread of estimations. It should be noted that the estimate that minimizes the LS cost function does not correspond to the true location, indicating that a higher number of delays alone does not guarantee accuracy. Finally, Figure 30 (d) using 27 delays (ES(27)) results in a single dominant cluster with only two estimates away from the cluster. Here, the most accurate localization estimate also minimizes the LS cost function, demonstrating that when nearly all relevant delays are included, the LS criterion becomes a reliable indicator of the true source position.

5.3 Partial conclusions

This chapter presented and evaluated the results of the proposed localization methods, based on TDOA measurements. The experiments demonstrated the viability of using acoustic signals for drone localization in noisy environments, validating both the TDOA-based LS approach and the NN method. It is worth mentioning that these cost functions were already validated during the DOA estimation experiments. While the LS-based techniques offer a lightweight and training-free solution, the NN approach outperformed them in terms of accuracy and robustness, particularly in reverberant conditions. The integration of reverberation fingerprints as input features in the NN model

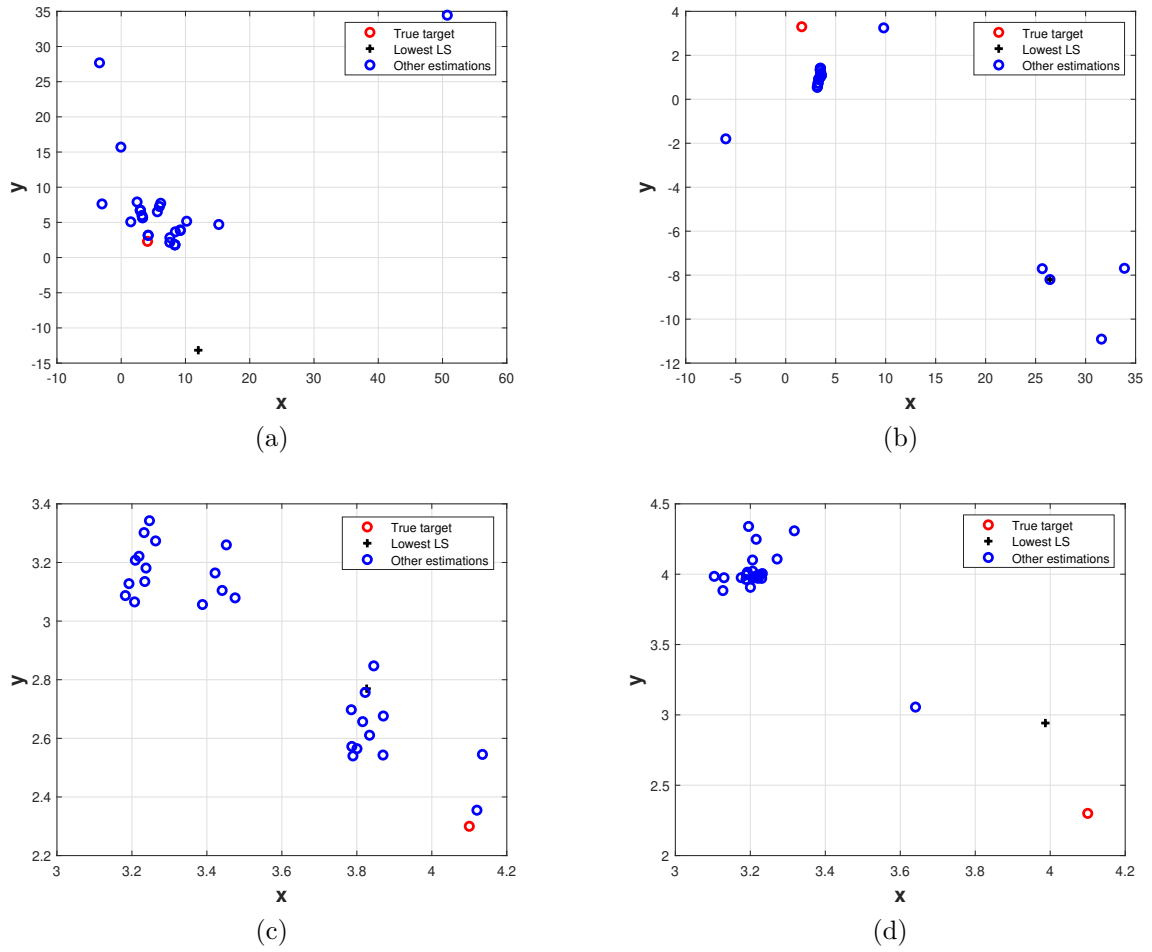


Figure 30 – Localization results obtained using different number, n , of actual delays $ES(n)$. (a) $ES(5)$; (b) $ES(6)$; (c) $ES(25)$; and (d) $ES(27)$.

proved effective in overcoming signal degradation caused by environmental noise and multipath effects. Overall, the results confirmed the capability of the proposed methods to estimate drone positions accurately and highlighted the trade-offs between complexity, adaptability, and performance in different operational contexts.

6 CONCLUSIONS

The proliferation of drones has introduced a pressing challenge for national defense and public safety, especially in scenarios where traditional detection technologies fall short due to environmental constraints such as noise, multipath propagation, and lack of line-of-sight. This thesis was developed under the pressing need to address this problem through the lens of acoustic signal processing, proposing methods capable of estimating the localization and DOA of drones operating in highly noisy and reverberant environments. With the growing interest in dual-use technologies, the work presented here aligns with the interests of defense and law enforcement, providing insights into how passive acoustic sensing can serve as a viable and effective counter-drone measure.

The general objective of this dissertation was to develop an acoustic-based methodology for estimating drone localization and DOA by analyzing the noise produced by their propellers. This objective was successfully achieved by proposing a novel framework that integrates ZCS and LS cost functions to identify accurate time delays even in the presence of multiple cross-correlation peaks.

Throughout this research, we have successfully implemented various algorithms to estimate the parameters of drones using acoustic signals, as detailed in Chapter 2. This comprehensive exploration has significantly advanced our understanding of drone parameter estimation and culminated in the development of some localization and DOA estimation algorithms that rival state-of-the-art methods.

Specific goals were also fulfilled, including creating a simulation framework to evaluate the impact of multiple correlation peaks, the experimental validation of acoustic TDOA measurements, and the proposal of methods that enhance estimation accuracy through exhaustive and heuristic searches. These goals were supported by detailed results, as discussed in Section 4.1 and Section 5.2, which confirm the effectiveness of the proposed methods in achieving accurate DOA and localization even under adverse conditions.

The findings presented in Chapter 4 underscore the potential of acoustics in the TDOA-based DOA estimation domain. Precise DOA estimation is paramount for the effective deployment of counter-drone measures. To address this, our research delved into the ZCS cost function alongside an additional LS cost function to augment the accuracy of our results. The insights gleaned from Chapter 4 shed light on the efficacy of our approach in estimating DOA accurately.

In addition to fulfilling its objectives, the thesis provided contributions beyond the originally stated goals. Among them, we highlight the results presented in Section 4.2, which prove: the capability of a low-complexity ZCS-based cost function to handle multiple

TDE candidates, the recognition of sub-optimality in ZCS, and the complementary role of LS for result refinement. The implementation of these techniques demonstrated the potential of combining classical signal processing with intelligent search strategies, offering a new perspective for real-time DOA estimation systems.

One of the key contributions of this thesis to the drone localization task lies in the evaluation of localization techniques using TDOA-based and learning-based methods. In particular, a detailed analysis demonstrated the performance differences between TDOA-based approaches and an NN model under realistic conditions. While traditional LS and its enhanced variants, including ZCS-LS, offered a baseline for localization based on acoustic time delays, the NN approach outperformed them in terms of accuracy in reverberant and noisy environments. This superior performance, achieved without relying solely on the primary cross-correlation peak, validates the integration of learned environmental features such as reverberation fingerprints. Despite the added complexity of requiring a training dataset, the NN model achieved lower localization errors and more consistent results. However, the NN requires a training dataset with fingerprints (many peaks at each possible position), and for a real-world application, this data may not always be available or may not be possible to collect samples due to a dangerous environment.

Looking ahead, several opportunities arise from the work developed in this thesis. One promising direction is the incorporation of distributed and synchronized microphone arrays for three-dimensional drone localization in wider and more complex scenarios by the application of Bearings-Only Target Motion Analysis [92, 143]. The implementation of robust time synchronization and data sharing protocols between sensor nodes will be essential for the real implementation of the localization algorithms without wired microphones. Future work may also explore the integration of other sensors to collect more features of the drones to enhance performance under variable conditions.

The utilization of distributed and synchronized devices will allow the estimation of drone localization in complex 3-D environments. These devices would collect drone acoustic signals and could initially employ low-cost hardware for feasibility testing. The main challenge lies in synchronizing these devices to facilitate accurate measurements of TDOA between pairs of microphones. This synchronization is important for ensuring the fidelity of the localization process. Concurrently, other endeavors will focus on recording the acoustic drone signals and sharing all signals to a processor unit to estimate the TDOA between pairs of signals and use them to locate the drones.

Ultimately, the culmination of this process will entail transmitting the derived localization data to a dedicated unit tasked with overseeing the situational awareness of the environment. This unit will serve as the nexus for interpreting and responding to the spatial dynamics of drone activity within the monitored airspace.

BIBLIOGRAPHY

- 1 MOHSAN, S. A. H. et al. Unmanned aerial vehicles (UAVs): practical aspects, applications, open challenges, security issues, and future trends. *Intelligent Service Robotics*, Springer, p. 1–29, 2023. Available at: <<https://doi.org/10.1007/s11370-022-00452-4>>.
- 2 AL-DOSARI, K. et al. A review of civilian drones systems, applications, benefits, safety, and security challenges. *The Effect of Information Technology on Business and Marketing Intelligence Systems*, Springer, p. 793–812, 2023. Available at: <https://doi.org/10.1007/978-3-031-12382-5_43>.
- 3 KUNERTOVA, D. The war in Ukraine shows the game-changing effect of drones depends on the game. *Bulletin of the Atomic Scientists*, Taylor & Francis, v. 79, n. 2, p. 95–102, 2023. Available at: <<https://doi.org/10.1080/00963402.2023.2178180>>.
- 4 ESKANDARIPOUR, H. et al. Last-mile drone delivery: Past, present, and future. *Drones*, MDPI, v. 7, n. 2, p. 77, 2023. Available at: <<https://doi.org/10.3390/drones7020077>>.
- 5 SERRENHO, F. G. et al. Gunshot airborne surveillance with rotary wing UAV-embedded microphone array. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 19, n. 19, p. 4271, 2019. Available at: <<https://doi.org/10.3390/s19194271>>.
- 6 AFSHAR-MOHAJER, N. et al. Use of a drone-based sensor as a field-ready technique for short-term concentration mapping of air pollutants: A modeling study. *Atmospheric Environment*, Elsevier, v. 294, p. 119476, 2023. Available at: <<https://doi.org/10.1016/j.atmosenv.2022.119476>>.
- 7 SHAWKY, M. et al. Traffic safety assessment for roundabout intersections using drone photography and conflict technique. *Ain Shams Engineering Journal*, Elsevier, p. 102115, 2023. Available at: <<https://doi.org/10.1016/j.asej.2023.102115>>.
- 8 GOODRICH, P. et al. Placement and drone flight path mapping of agricultural soil sensors using machine learning. *Computers and Electronics in Agriculture*, Elsevier, v. 205, p. 107591, 2023. Available at: <<https://doi.org/10.1016/j.compag.2022.107591>>.
- 9 PLEDGER, T. The role of drones in future terrorist attacks. *Association of the United States Army*, v. 26, 2021. Available at: <<https://www.ausa.org/publications/role-drones-future-terrorist-attacks>>.
- 10 SUOJANEN, M. et al. An example of scenario-based evaluation of military capability areas. An impact assessment of alternative systems on operations. In: IEEE. *Annual IEEE Systems Conference (SysCon) Proceedings*. 2015. p. 601–607. Available at: <<https://doi.org/10.1109/SYSCON.2015.7116817>>.
- 11 WATLING, J. et al. Achieving lethal effects by small unmanned aerial vehicles: Opportunities and limitations. *The RUSI Journal*, Taylor & Francis, v. 164, n. 1, p. 40–51, 2019. Available at: <<https://doi.org/10.1080/03071847.2019.1605017>>.

- 12 KREPS, S. et al. Drones in modern war: evolutionary, revolutionary, or both? *Defense & Security Analysis*, Taylor & Francis, p. 1–4, 2023. Available at: <<https://doi.org/10.1080/14751798.2023.2178599>>.
- 13 NASEEM, A. et al. Decision support system for optimum decision making process in threat evaluation and weapon assignment: Current status, challenges and future directions. *Annual Reviews in Control*, Elsevier, v. 43, p. 169–187, 2017. Available at: <<https://doi.org/10.1016/j.arcontrol.2017.03.003>>.
- 14 ANWAR, M. Z. et al. Machine learning inspired sound-based amateur drone detection for public safety applications. *IEEE Transactions on Vehicular Technology*, IEEE, v. 68, n. 3, p. 2526–2534, 2019. Available at: <<https://doi.org/10.1109/TVT.2019.2893615>>.
- 15 AKBAL, E. et al. An automated accurate sound-based amateur drone detection method based on skinny pattern. *Digital Signal Processing*, Elsevier, p. 104012, 2023. Available at: <<https://doi.org/10.1016/j.dsp.2023.104012>>.
- 16 AL-EMADI, S. et al. Audio-based drone detection and identification using deep learning techniques with dataset enhancement through generative adversarial networks. *Sensors*, MDPI, v. 21, n. 15, p. 4953, 2021. Available at: <<https://doi.org/10.1016/j.dsp.2023.104012>>.
- 17 KYRITSIS, A. et al. Small UAS online audio DOA estimation and real-time identification using machine learning. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 22, n. 22, p. 8659, 2022. Available at: <<https://doi.org/10.3390/s22228659>>.
- 18 SHI, Z. et al. An acoustic-based surveillance system for amateur drones detection and localization. *IEEE Transactions on Vehicular Technology*, IEEE, v. 69, n. 3, p. 2731–2739, 2020. Available at: <<https://doi.org/10.1109/TVT.2020.2964110>>.
- 19 YANG, C. et al. DOA estimation using amateur drones harmonic acoustic signals. In: IEEE. *2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*. 2018. p. 587–591. Available at: <<https://doi.org/10.1109/SAM.2018.8448797>>.
- 20 QIU, W. et al. Acoustic slam based on the direction-of-arrival and the direct-to-reverberant energy ratio. *Drones*, MDPI, v. 7, n. 2, p. 120, 2023. Available at: <<https://doi.org/10.3390/drones7020120>>.
- 21 WU, Z. et al. A drone localization system based on multiple time delays fusion. In: IEEE. *2019 IEEE Radar Conference (RadarConf)*. 2019. p. 1–6. Available at: <<https://doi.org/10.1109/RADAR.2019.8835531>>.
- 22 FERNANDES, R. P. et al. Applying the majority voting rule in acoustic detection and classification of drones. In: . Sociedade Brasileira de Telecomunicações, 2021. Available at: <<https://doi.org/10.14209/sbrt.2021.1570726624>>.
- 23 IBRAHIM, O. A. et al. Noise2Weight: On detecting payload weight from drones acoustic emissions. *Future Generation Computer Systems*, Elsevier, v. 134, p. 319–333, 2022. Available at: <<https://doi.org/10.1016/j.future.2022.03.041>>.
- 24 DALE, H. et al. SNR-dependent drone classification using convolutional neural networks. *IET Radar, Sonar & Navigation*, Wiley Online Library, v. 16, n. 1, p. 22–33, 2022. Available at: <<https://doi.org/10.1049/rsn2.12161>>.

- 25 AL-SA'D, M. F. et al. RF-based drone detection and identification using deep learning approaches: An initiative towards a large open source drone database. *Future Generation Computer Systems*, Elsevier, v. 100, p. 86–97, 2019. Available at: <<https://doi.org/10.1016/j.future.2019.05.007>>.
- 26 SEIDALIYEVA, U. et al. Deep residual neural network-based classification of loaded and unloaded UAV images. In: IEEE. *2020 Fourth IEEE International Conference on Robotic Computing (IRC)*. 2020. p. 465–469. Available at: <<https://doi.org/10.1109/IRC.2020.00088>>.
- 27 LEE, D. et al. Drone detection and identification system using artificial intelligence. In: IEEE. *2018 International Conference on Information and Communication Technology Convergence (ICTC)*. 2018. p. 1131–1133. Available at: <<https://doi.org/10.1109/ICTC.2018.8539442>>.
- 28 LAI, Y.-C. et al. Detection of a moving UAV based on deep learning-based distance estimation. *Remote Sensing*, MDPI, v. 12, n. 18, p. 3035, 2020. Available at: <<https://doi.org/10.3390/rs12183035>>.
- 29 XIE, J. et al. Adaptive switching spatial-temporal fusion detection for remote flying drones. *IEEE Transactions on Vehicular Technology*, IEEE, v. 69, n. 7, p. 6964–6976, 2020. Available at: <<https://doi.org/10.1109/TVT.2020.2993863>>.
- 30 SEDUNOV, A. et al. UAV passive acoustic detection. In: IEEE. *2018 IEEE International Symposium on Technologies for Homeland Security (HST)*. 2018. p. 1–6. Available at: <<https://doi.org/10.1109/THS.2018.8574129>>.
- 31 LIU, H. et al. Acoustic source localization for anti-UAV based on machine learning in wireless sensor networks. In: IEEE. *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*. 2020. p. 1142–1147. Available at: <<https://doi.org/10.1109/ICIEA48937.2020.9248139>>.
- 32 CHRISTNACHER, F. et al. Optical and acoustical UAV detection. In: SPIE. *Electro-Optical Remote Sensing X*. 2016. v. 9988, p. 83–95. Available at: <<https://doi.org/10.1117/12.2240752>>.
- 33 TOMA, A. et al. CNN-based processing of radio frequency signals for augmenting acoustic source localization and enhancement in UAV security applications. In: IEEE. *2021 International Conference on Military Communication and Information Systems (ICMCIS)*. 2021. p. 1–5. Available at: <<https://doi.org/10.1109/ICMCIS52405.2021.9486424>>.
- 34 TOMA, A. et al. Onboard audio and video processing for secure detection, localization, and tracking in counter-UAV applications. *Procedia Computer Science*, Elsevier, v. 205, p. 20–27, 2022. Available at: <<https://doi.org/10.1016/j.procs.2022.09.003>>.
- 35 TAHA, B. et al. Machine learning-based drone detection and classification: State-of-the-art in research. *IEEE Access*, IEEE, v. 7, p. 138669–138682, 2019. Available at: <<https://doi.org/10.1109/ACCESS.2019.2942944>>.
- 36 LYKOU, G. et al. Defending airports from UAS: A survey on cyber-attacks and counter-drone sensing technologies. *Sensors*, MDPI, v. 20, n. 12, p. 3537, 2020. Available at: <<https://doi.org/10.3390/s20123537>>.

- 37 UTEBAYEVA, D. et al. Practical study of recurrent neural networks for efficient real-time drone sound detection: A review. *Drones*, MDPI, v. 7, n. 1, p. 26, 2022. Available at: <<https://doi.org/10.3390/drones7010026>>.
- 38 WANG, L. et al. Ear in the sky: Ego-noise reduction for auditory micro aerial vehicles. In: IEEE. *2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. 2016. p. 152–158. Available at: <<https://doi.org/10.1109/AVSS.2016.7738063>>.
- 39 WANG, L. et al. Acoustic sensing from a multi-rotor drone. *IEEE Sensors Journal*, IEEE, v. 18, n. 11, p. 4570–4582, 2018. Available at: <<https://doi.org/10.1109/JSEN.2018.2825879>>.
- 40 MUKHUTDINOV, D. et al. Deep learning models for single-channel speech enhancement on drones. *IEEE Access*, IEEE, 2023. Available at: <<https://doi.org/10.1109/ACCESS.2023.3253719>>.
- 41 FARAJI, M. M. et al. Sound source localization in wide-range outdoor environment using distributed sensor network. *IEEE Sensors Journal*, IEEE, v. 20, n. 4, p. 2234–2246, 2019. Available at: <<https://doi.org/10.1109/JSEN.2019.2950447>>.
- 42 CHÁVEZ, K. et al. Emulating underdogs: Tactical drones in the Russia-Ukraine war. *Contemporary Security Policy*, Taylor & Francis, v. 44, n. 4, p. 592–605, 2023. Available at: <<https://doi.org/10.1080/13523260.2023.2257964>>.
- 43 DEVORE, M. R. “no end of a lesson:” observations from the first high-intensity drone war. *Defense & Security Analysis*, Taylor & Francis, v. 39, n. 2, p. 263–266, 2023. Available at: <<https://doi.org/10.1080/14751798.2023.2178571>>.
- 44 BREWCZYŃSKI, K. D. et al. Methods for assessing the effectiveness of modern counter unmanned aircraft systems. *Remote Sensing*, MDPI, v. 16, n. 19, p. 3714, 2024. Available at: <<https://doi.org/10.3390/rs16193714>>.
- 45 CHAUHAN, D. et al. Nation’s defense: A comprehensive review of anti-drone systems and strategies. *IEEE Access*, IEEE, 2025. Available at: <<https://doi.org/10.1109/ACCESS.2025.3550338>>.
- 46 YANG, T. et al. Fast and robust super-resolution doa estimation for UAV swarms. *Signal Processing*, Elsevier, v. 188, p. 108187, 2021. Available at: <<https://doi.org/10.1016/j.sigpro.2021.108187>>.
- 47 CHANG, X. et al. A surveillance system for drone localization and tracking using acoustic arrays. In: IEEE. *2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*. 2018. p. 573–577. Available at: <<https://doi.org/10.1109/SAM.2018.8448409>>.
- 48 SUN, Y. et al. Aim: Acoustic inertial measurement for indoor drone localization and tracking. In: *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*. [s.n.], 2022. p. 476–488. Available at: <<https://doi.org/10.1145/3560905.3568499>>.
- 49 YANG, F. et al. Practical investigation of a MIMO radar system capabilities for small drones detection. *IET Radar, Sonar & Navigation*, Wiley Online Library, v. 15, n. 7, p. 760–774, 2021.

- 50 GRIECO, G. et al. Detection, tracking, and identification of drones: an overview on counter-UAS techniques, and open challenges. In: IEEE. *2025 Integrated Communications, Navigation and Surveillance Conference (ICNS)*. 2025. p. 1–8. Available at: <<https://doi.org/10.1109/ICNS65417.2025.10976862>>.
- 51 FERNANDES, R. P. et al. Enhancing TDE-based drone DoA estimation with genetic algorithms and zero cyclic sum. In: *XVI Brazilian Conference on Computational Intelligence*. Congresso Brasileiro de Inteligência Computacional, 2023. Available at: <<http://dx.doi.org/10.21528/CBIC2023-115>>.
- 52 DEFESA, M. da. *Estratégia Nacional de Defesa*. 2024. <https://www.gov.br/defesa/pt-br/arquivos/ajuste-01/estado_e_defesa/pnd_end_congresso_.pdf>. [Online; accessed 14-november-2024].
- 53 FERNANDES, R. P. *Drone localization and DOA estimation*. 2025. <<https://github.com/RigelFernandes/drone-localization-DOA-estimation>>. [Online; accessed 14-July-2025].
- 54 KNAPP, C. et al. The generalized correlation method for estimation of time delay. *IEEE transactions on acoustics, speech, and signal processing*, IEEE, v. 24, n. 4, p. 320–327, 1976. Available at: <<https://doi.org/10.1109/TASSP.1976.1162830>>.
- 55 HALL, M. A combinatorial problem on abelian groups. *Proceedings of the American Mathematical Society*, JSTOR, v. 3, n. 4, p. 584–587, 1952. Available at: <<https://doi.org/10.2307/2032592>>.
- 56 BORZINO, A. M. C. R. et al. Consistent DOA estimation of heavily noisy gunshot signals using a microphone array. *IET Radar, Sonar & Navigation*, Wiley Online Library, v. 10, n. 9, p. 1519–1527, 2016. Available at: <<https://doi.org/10.1049/iet-rsn.2016.0015>>.
- 57 BOSWORTH, B. T. et al. Estimating signal-to-noise ratio (snr). *IEEE Journal of Oceanic Engineering*, IEEE, v. 33, n. 4, p. 414–418, 2008. Available at: <<https://doi.org/10.1109/JOE.2008.2001780>>.
- 58 DANIEL, J. et al. Time domain velocity vector for retracing the multipath propagation. In: IEEE. *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 2020. p. 421–425. Available at: <<https://doi.org/10.1109/ICASSP40776.2020.9054561>>.
- 59 PLINGE, A. et al. Acoustic microphone geometry calibration: An overview and experimental evaluation of state-of-the-art algorithms. *IEEE Signal Processing Magazine*, IEEE, v. 33, n. 4, p. 14–29, 2016. Available at: <<https://doi.org/10.1109/MSP.2016.2555198>>.
- 60 SMITH, J. O. et al. Closed-form least-squares source location estimation from range-difference measurements. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 35, n. 12, p. 1661–1669, Dec. 1987. ISSN 0096-3518. Available at: <<https://doi.org/10.1109/TASSP.1987.1165089>>.
- 61 KHALAF-ALLAH, M. An extended closed-form least-squares solution for three-dimensional hyperbolic geolocation. In: *2014 IEEE Symposium on Industrial Electronics Applications (ISIEA)*. [s.n.], 2014. p. 7–11. Available at: <<https://doi.org/10.1109/ISIEA.2014.8049862>>.

- 62 FREIRE, I. L. Robust direction-of-arrival by matched-lags, applied to gunshots. *The Journal of the Acoustical Society of America*, AIP Publishing, v. 135, n. 6, p. EL246–EL251, 2014. Available at: <<https://doi.org/10.1121/1.4874223>>.
- 63 SOUSA, M. N. de et al. Enhancement of localization systems in NLOS urban scenario with multipath ray tracing fingerprints and machine learning. *Sensors*, MDPI, v. 18, n. 11, 2018. ISSN 1424-8220. Available at: <<https://doi.org/10.3390/s18114073>>.
- 64 SOUSA, M. N. de et al. Improving the performance of a radio-frequency localization system in adverse outdoor applications. *EURASIP Journal on Wireless Communications and Networking*, v. 2021, 2021. ISSN 1424-8220. Available at: <<https://doi.org/10.1186/s13638-021-02001-6>>.
- 65 APOLINÁRIO JR., J. A. et al. Exploiting reverberation fingerprint for a neural network based acoustic emitter localization. In: IEEE. *2024 10th International Conference on Control, Decision and Information Technologies (CoDIT)*. Valletta, Malta, 2024. p. 1–6. Available at: <<http://doi.org/10.1109/CoDIT62066.2024.10708420>>.
- 66 KHAN, M. A. et al. On the detection of unauthorized drones—techniques and future perspectives: A review. *IEEE Sensors Journal*, IEEE, v. 22, n. 12, p. 11439–11455, 2022. Available at: <<https://doi.org/10.1109/JSEN.2022.3171293>>.
- 67 KADYROV, D. et al. Improvements to the Stevens drone acoustic detection system. In: AIP PUBLISHING. *Proceedings of Meetings on Acoustics*. 2022. v. 46, n. 1. Available at: <<https://doi.org/10.1121/2.0001602>>.
- 68 AN, W. et al. Estimation of number of unmanned aerial vehicles in a scene utilizing acoustic signatures and machine learning. *The Journal of the Acoustical Society of America*, AIP Publishing, v. 154, n. 1, p. 533–546, 2023. Available at: <<https://doi.org/10.1121/10.0020292>>.
- 69 FARIA, M. M. D. et al. Analyzing its governance initiatives with game theory: A systematic literature review. *J. Softw.*, v. 10, n. 9, p. 1056–1069, 2015. Available at: <<http://doi.org/10.17706/jsw.10.9.1056-1069>>.
- 70 RAHMAN, M. H. et al. A comprehensive survey of unmanned aerial vehicles detection and classification using machine learning approach: Challenges, solutions, and future directions. *Remote Sensing*, Multidisciplinary Digital Publishing Institute, v. 16, n. 5, p. 879, 2024. Available at: <<https://doi.org/10.3390/rs16050879>>.
- 71 SEIDALIYEVA, U. et al. Advances and challenges in drone detection and classification techniques: A state-of-the-art review. *Sensors*, MDPI, v. 24, n. 1, p. 125, 2023. Available at: <<https://doi.org/10.3390/s24010125>>.
- 72 TEJERA-BERENGUE, D. et al. Analysis of distance and environmental impact on UAV acoustic detection. *Electronics*, MDPI, v. 13, n. 3, p. 643, 2024. Available at: <<https://doi.org/10.3390/electronics13030643>>.
- 73 VALLIAPPAN, N. H. et al. Enhancing gun detection with transfer learning and YAMNet audio classification. *IEEE Access*, IEEE, 2024. Available at: <<https://doi.org/10.1109/ACCESS.2024.3392649>>.

- 74 TEJERA-BERENGUE, D. et al. Acoustic-based detection of UAVs using machine learning: Analysis of distance and environmental effects. In: IEEE. *2023 IEEE Sensors Applications Symposium (SAS)*. 2023. p. 1–6. Available at: <<https://doi.org/10.1109/SAS58821.2023.10254127>>.
- 75 MARTINEZ-CARRANZA, J. et al. A review on auditory perception for unmanned aerial vehicles. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 20, n. 24, p. 7276, 2020. Available at: <<https://doi.org/10.3390/s20247276>>.
- 76 HARVEY, B. et al. Acoustic detection of a fixed-wing UAV. *Drones*, Multidisciplinary Digital Publishing Institute, v. 2, n. 1, p. 4, 2018. Available at: <<https://doi.org/10.3390/drones2010004>>.
- 77 WANG, L. et al. A blind source separation framework for ego-noise reduction on multi-rotor drones. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, IEEE, v. 28, p. 2523–2537, 2020. Available at: <<https://doi.org/10.1109/TASLP.2020.3015027>>.
- 78 WANG, L. et al. Deep learning assisted time-frequency processing for speech enhancement on drones. *IEEE Transactions on Emerging Topics in Computational Intelligence*, IEEE, 2020. Available at: <<https://doi.org/10.1109/TETCI.2020.3014934>>.
- 79 DING, S. et al. Drone detection and tracking system based on fused acoustical and optical approaches. *Advanced Intelligent Systems*, Wiley Online Library, v. 5, n. 10, p. 2300251, 2023. Available at: <<https://doi.org/10.1002/aisy.202300251>>.
- 80 DOSTER, H. G. UAV payload identification with acoustic emissions and cell phone devices. 2022. Available at: <<https://apps.dtic.mil/sti/citations/AD1166859>>.
- 81 DOSTER, H. et al. UAV payload identification with acoustic emissions and cell phones. In: *International Conference on Cyber Warfare and Security*. [s.n.], 2023. v. 18, n. 1, p. 523–533. Available at: <<https://doi.org/10.34190/iccws.18.1.956>>.
- 82 BORZINO, A. M. et al. Robust DOA estimation of heavily noisy gunshot signals. In: IEEE. *2015 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 2015. p. 449–453. Available at: <<https://doi.org/10.1109/ICASSP.2015.7178009>>.
- 83 FIROOZABADI, A. D. et al. 3d multiple sound source localization by proposed T-shaped circular distributed microphone arrays in combination with GEVD and adaptive GCC-PHAT/ML algorithms. *Sensors*, MDPI, v. 22, n. 3, p. 1011, 2022. Available at: <<https://doi.org/10.3390/s22031011>>.
- 84 BU, S. et al. TDOA estimation of speech source in noisy reverberant environments. In: IEEE. *2022 IEEE Spoken Language Technology Workshop (SLT)*. 2023. p. 1059–1066. Available at: <<https://doi.org/10.1109/SLT54892.2023.10023256>>.
- 85 WANG, Z.-Q. et al. Robust TDOA estimation based on time-frequency masking and deep neural networks. In: *Interspeech*. [s.n.], 2018. p. 322–326. Available at: <<https://doi.org/10.21437/Interspeech.2018-1652>>.
- 86 LIAQUAT, M. U. et al. Localization of sound sources: A systematic review. *Energies*, MDPI, v. 14, n. 13, p. 3910, 2021. Available at: <<https://doi.org/10.3390/en14133910>>.

- 87 FREIRE, I. L. et al. GCC-based DoA estimation of overlapping muzzleblast and shockwave components of gunshot signals. In: IEEE. *2011 IEEE Second Latin American Symposium on Circuits and Systems (LASCAS)*. 2011. p. 1–4. Available at: <<https://doi.org/10.1109/LASCAS.2011.5750273>>.
- 88 CALDERON, D. M. P. et al. Shooter localization based on DoA estimation of gunshot signals and digital map information. *IEEE Latin America Transactions*, IEEE, v. 13, n. 2, p. 441–447, 2015. Available at: <<https://doi.org/10.1109/TLA.2015.7055562>>.
- 89 VARMA, K. et al. Robust tde-based doa estimation for compact audio arrays. In: IEEE. *Sensor Array and Multichannel Signal Processing Workshop Proceedings, 2002*. 2002. p. 214–218. Available at: <<https://doi.org/10.1109/SAM.2002.1191031>>.
- 90 SCHEUING, J. et al. Disambiguation of TDOA estimation for multiple sources in reverberant environments. *IEEE Transactions on Audio, Speech, and Language Processing*, IEEE, v. 16, n. 8, p. 1479–1489, 2008. Available at: <<https://doi.org/10.1109/TASL.2008.2004533>>.
- 91 FERNANDES, R. P. et al. Investigating the potential of UAV for gunshot DoA estimation and shooter localization. In: SBRT. *Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*. 2016. p. 383–387. Available at: <<http://dx.doi.org/10.14209/sbrt.2016.52>>.
- 92 FERNANDES, R. P. et al. Bearings-only aerial shooter localization using a microphone array mounted on a drone. In: IEEE. *2017 IEEE 8th Latin American Symposium on Circuits & Systems (LASCAS)*. 2017. p. 1–4. Available at: <<https://doi.org/10.1109/LASCAS.2017.7948081>>.
- 93 FERNANDES, R. P. et al. Airborne DoA estimation of gunshot acoustic signals using drones with application to sniper localization systems. In: SPIE. *Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security, Defense, and Law Enforcement Applications XVI*. 2017. v. 10184, p. 51–57. Available at: <<https://doi.org/10.1117/12.2262782>>.
- 94 JENSEN, J. R. et al. DOA estimation of audio sources in reverberant environments. In: IEEE. *2016 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 2016. p. 176–180. Available at: <<https://doi.org/10.1109/ICASSP.2016.7471660>>.
- 95 DRÉMEAU, A. et al. DOA estimation in structured phase-noisy environments. In: IEEE. *2017 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 2017. p. 3176–3180. Available at: <<https://doi.org/10.1109/ICASSP.2017.7952742>>.
- 96 CUI, X. et al. Approximate closed-form TDOA-based estimator for acoustic direction finding via constrained optimization. *IEEE Sensors Journal*, IEEE, v. 18, n. 8, p. 3360–3371, 2018. Available at: <<https://doi.org/10.1109/JSEN.2018.2803150>>.
- 97 EVERS, C. et al. DoA reliability for distributed acoustic tracking. *IEEE Signal Processing Letters*, IEEE, v. 25, n. 9, p. 1320–1324, 2018. Available at: <<https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=97>>.

- 98 BAN, Y. et al. Tracking multiple audio sources with the von Mises distribution and variational EM. *IEEE Signal Processing Letters*, IEEE, v. 26, n. 6, p. 798–802, 2019. Available at: <<https://doi.org/10.1109/LSP.2019.2908376>>.
- 99 SEWTZ, M. et al. Robust MUSIC-Based sound source localization in reverberant and echoic environments. In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. [s.n.], 2020. p. 2474–2480. Available at: <<https://doi.org/10.1109/IROS45743.2020.9340826>>.
- 100 SUN, B. et al. Direction-of-arrival estimation of acoustic sources using acoustic array based on SOM and BP Neural Network. In: *Proceedings of the 6th International Conference on Digital Signal Processing*. [s.n.], 2022. p. 205–216. Available at: <<https://doi.org/10.1145/3529570.3529605>>.
- 101 SHUJAU, M. et al. Designing acoustic vector sensors for localisation of sound sources in air. In: IEEE. *2009 17th European Signal Processing Conference*. 2009. p. 849–853. Available at: <<https://ieeexplore.ieee.org/abstract/document/7077457>>.
- 102 CHUNG, M.-A. et al. Sound localization based on acoustic source using multiple microphone array in an indoor environment. *Electronics*, MDPI, v. 11, n. 6, p. 1–12, 2022. Available at: <<https://doi.org/10.3390/electronics11060890>>.
- 103 KAHRS, M. et al. (Ed.). *Applications of Digital Signal Processing to Audio and Acoustics*. Kluwer Academic Publishers, 2002. Available at: <<https://doi.org/10.1007/b117882>>.
- 104 ASHRAF, I. et al. *Recent Advancements in Indoor Positioning and Localization*. MDPI, 2022. 1–5 p. Available at: <<https://doi.org/10.3390/electronics11132047>>.
- 105 RIBEIRO, F. et al. Using reverberation to improve range and elevation discrimination for small array sound source localization. *IEEE Transactions on Audio, Speech, and Language Processing*, v. 18, n. 7, p. 1781–1792, 2010. Available at: <<https://doi.org/10.1109/TASL.2010.2052250>>.
- 106 YAPAR, Ç. et al. Locunet: Fast urban positioning using radio maps and deep learning. In: *2022 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. [s.n.], 2022. p. 4063–4067. Available at: <<https://doi.org/10.1109/ICASSP43922.2022.9747240>>.
- 107 LICITRA, G. et al. Acoustic beamforming algorithms and their applications in environmental noise. *Current Pollution Reports*, Springer, p. 1–24, 2023. Available at: <<https://doi.org/10.1007/s40726-023-00264-9>>.
- 108 SIJTSMA, P. Clean based on spatial source coherence. *International journal of aeroacoustics*, SAGE Publications Sage UK: London, England, v. 6, n. 4, p. 357–374, 2007. Available at: <<https://doi.org/10.1260/147547207783359459>>.
- 109 BROOKS, T. F. et al. A deconvolution approach for the mapping of acoustic sources (DAMAS) determined from phased microphone arrays. *Journal of sound and vibration*, Elsevier, v. 294, n. 4-5, p. 856–879, 2006. Available at: <<https://doi.org/10.1016/j.jsv.2005.12.046>>.

- 110 SCHMIDT, R. Multiple emitter location and signal parameter estimation. *IEEE transactions on antennas and propagation*, IEEE, v. 34, n. 3, p. 276–280, 1986. Available at: <<https://doi.org/10.1109/TAP.1986.1143830>>.
- 111 RAMOS, A. L. et al. Delay-and-sum beamforming for direction of arrival estimation applied to gunshot acoustics. In: SPIE. *Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense X*. 2011. v. 8019, p. 162–170. Available at: <<https://doi.org/10.1117/12.886833>>.
- 112 CHIARIOTTI, P. et al. Acoustic beamforming for noise source localization—reviews, methodology and applications. *Mechanical Systems and Signal Processing*, Elsevier, v. 120, p. 422–448, 2019. Available at: <<https://doi.org/10.1016/j.ymssp.2018.09.019>>.
- 113 YANG, Y. et al. Functional delay and sum beamforming for three-dimensional acoustic source identification with solid spherical arrays. *Journal of Sound and Vibration*, Elsevier, v. 373, p. 340–359, 2016. Available at: <<https://doi.org/10.1016/j.jsv.2016.03.024>>.
- 114 LEE, J.-Y. et al. Acoustic DOA estimation: An approximate maximum likelihood approach. *IEEE systems journal*, IEEE, v. 8, n. 1, p. 131–141, 2013. Available at: <<https://doi.org/10.1109/JSYST.2013.2260630>>.
- 115 STOICA, P. et al. Music, maximum likelihood, and Cramer-Rao bound. *IEEE Transactions on Acoustics, speech, and signal processing*, IEEE, v. 37, n. 5, p. 720–741, 1989. Available at: <<https://doi.org/10.1109/29.17564>>.
- 116 HUANG, G. et al. Direction-of-arrival estimation of passive acoustic sources in reverberant environments based on the householder transformation. *The Journal of the Acoustical Society of America*, AIP Publishing, v. 138, n. 5, p. 3053–3060, 2015. Available at: <<https://doi.org/10.1121/1.4934954>>.
- 117 WAJID, M. et al. Direction-of-arrival estimation algorithms using single acoustic vector-sensor. In: IEEE. *2017 International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT)*. 2017. p. 84–88. Available at: <<https://doi.org/10.1109/MSPCT.2017.8363979>>.
- 118 KOTUS, J. et al. Calibration of acoustic vector sensor based on MEMS microphones for DOA estimation. *Applied Acoustics*, Elsevier, v. 141, p. 307–321, 2018. Available at: <<https://doi.org/10.1016/j.apacoust.2018.07.025>>.
- 119 HU, Y. et al. Direction of arrival estimation of multiple acoustic sources using a maximum likelihood method in the spherical harmonic domain. *Applied Acoustics*, Elsevier, v. 135, p. 85–90, 2018. Available at: <<https://doi.org/10.1016/j.apacoust.2018.02.005>>.
- 120 SRINIVAS, M. et al. Genetic algorithms: A survey. *computer*, IEEE, v. 27, n. 6, p. 17–26, 1994. Available at: <<https://doi.org/10.1109/2.294849>>.
- 121 KASSIR, H. A. et al. A review of the state-of-the-art and future challenges of deep learning-based beamforming. *IEEE Access*, IEEE, 2022. Available at: <<https://doi.org/10.1109/ACCESS.2022.3195299>>.
- 122 XIAO, Y. et al. High-resolution acoustic beamforming based on genetic algorithms. *Mechanical Systems and Signal Processing*, Elsevier, v. 204, p. 110840, 2023. Available at: <<https://doi.org/10.1016/j.ymssp.2023.110840>>.

- 123 XIAO, X. et al. A learning-based approach to direction of arrival estimation in noisy and reverberant environments. In: IEEE. *2015 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 2015. p. 2814–2818. Available at: <<https://doi.org/10.1109/ICASSP.2015.7178484>>.
- 124 CHAKRABARTY, S. et al. Broadband DOA estimation using convolutional neural networks trained with noise signals. In: IEEE. *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. 2017. p. 136–140. Available at: <<https://doi.org/10.1109/WASPAA.2017.8170010>>.
- 125 BLANCHARD, T. et al. Acoustic localization and tracking of a multi-rotor unmanned aerial vehicle using an array with few microphones. *The Journal of the Acoustical Society of America*, AIP Publishing, v. 148, n. 3, p. 1456–1467, 2020. Available at: <<https://doi.org/10.1121/10.0001930>>.
- 126 ITARE, N. et al. Acoustic estimation of the direction of arrival of an unmanned aerial vehicle based on frequency tracking in the time-frequency plane. *Sensors*, MDPI, v. 22, n. 11, p. 4021, 2022. Available at: <<https://doi.org/10.3390/s22114021>>.
- 127 SUN, Y. et al. Indoor drone localization and tracking based on acoustic inertial measurement. *IEEE Transactions on Mobile Computing*, IEEE, 2023. Available at: <<https://doi.org/10.1109/TMC.2023.3335860>>.
- 128 SEDUNOV, A. et al. Stevens drone detection acoustic system and experiments in acoustics uav tracking. In: IEEE. *2019 IEEE International Symposium on Technologies for Homeland Security (HST)*. 2019. p. 1–7. Available at: <<https://doi.org/10.1109/HST47167.2019.9032916>>.
- 129 CHERVONIAK, Y. et al. TDoA and Doppler shift estimation method for passive acoustic location of flying vehicles. In: IEEE. *2018 IEEE 17th International Conference on Mathematical Methods in Electromagnetic Theory (MMET)*. 2018. p. 119–122. Available at: <<https://doi.org/10.1109/MMET.2018.8460359>>.
- 130 WU, S. et al. Sound source localization for unmanned aerial vehicles in low signal-to-noise ratio environments. *Remote Sensing*, MDPI, v. 16, n. 11, p. 1847, 2024. Available at: <<https://doi.org/10.3390/rs16111847>>.
- 131 LAUZON, J.-S. et al. Localization of RW-UAVs using particle filtering over distributed microphone arrays. In: IEEE. *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2017. p. 2479–2484. Available at: <<https://doi.org/10.1109/IROS.2017.8206065>>.
- 132 BAGGENSTOSS, P. M. et al. Efficient phase-based acoustic tracking of drones using a microphone array. In: IEEE. *2019 27th European Signal Processing Conference (EUSIPCO)*. 2019. p. 1–5. Available at: <<https://doi.org/10.23919/EUSIPCO.2019.8902972>>.
- 133 CHEN, T. et al. Research on a sound source localization method for uav detection based on improved empirical mode decomposition. *Sensors*, MDPI, v. 24, n. 9, p. 2701, 2024. Available at: <<https://doi.org/10.3390/s24092701>>.
- 134 FERNANDES, R. P. et al. A reduced complexity acoustic-based 3D DoA estimation with zero cyclic sum. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 24, n. 7, p. 2344, 2024. Available at: <<https://doi.org/10.3390/s24072344>>.

- 135 MINIDSP. *UMA-8 USB mic array - V2.0*. 2023. <<https://www.minidsp.com/products/usb-audio-interface/uma-8-microphone-array/>>. [Online; accessed 05-june-2023].
- 136 MARSANO-CORNEJO, M.-J. et al. Comparison of the acoustic parameters obtained with different smartphones and a professional microphone. *Acta Otorrinolaringologica (English Edition)*, Elsevier, v. 73, n. 1, p. 51–55, 2022. Available at: <<https://doi.org/10.1016/j.otoeng.2020.08.009>>.
- 137 TIMANÁ, L. C. R. et al. Technical feasibility for the mobile measurement of noise pollution by remotely piloted aircraft system. In: SPRINGER. *International Conference on Applied Technologies*. 2019. p. 219–230. Available at: <https://doi.org/10.1007/978-3-030-42531-9_18>.
- 138 SMITH, J. O. et al. The spherical interpolation method of source localization. *IEEE Journal of Oceanic Engineering*, OE-12, n. 1, p. 246–252, Jan. 1987. ISSN 0364-9059. Available at: <<https://doi.org/10.1109/JOE.1987.1145217>>.
- 139 STOICA, P. et al. Lecture notes – Source localization from range-difference measurements. *IEEE Signal Processing Magazine*, v. 23, n. 6, p. 63–66, Nov. 2006. ISSN 1053-5888. Available at: <<https://doi.org/10.1109/SP-M.2006.248717>>.
- 140 KNAPP, C. H. et al. The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 24, n. 4, p. 320–327, Aug. 1976. ISSN 0096-3518. Available at: <<https://doi.org/10.1109/TASSP.1976.1162830>>.
- 141 APOLINÁRIO JR., J. A. et al. A Data-selective LS Solution to TDOA-based Source Localization. In: *2019 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. [s.n.], 2019. p. 4400–4404. Available at: <<https://doi.org/10.1109/ICASSP.2019.8682664>>.
- 142 YOUNG, R. W. Sabine reverberation equation revisited. *The Journal of the Acoustical Society of America*, v. 91, n. 4, 1992. Available at: <<https://doi.org/10.1121/1.403389>>.
- 143 FERNANDES, R. P. *Análise do movimento de alvos a partir de sinais captados por um VANT*. Dissertation (Master's in Defense Engineering) — Instituto Militar de Engenharia, Rio de Janeiro, RJ, Brazil, 2017. Available at: <<http://bdex.eb.mil.br/jspui/handle/123456789/9096>>.

APPENDIX A – PUBLISHED WORKS

List of the conference and journal papers produced during the doctoral course:

Conference Paper:

1. FERNANDES, R. P.; APOLINÁRIO JR., J. A. **Underwater target classification with optimized feature selection based on genetic algorithms**. In: SBRT. Simpósio Brasileiro de Telecomunicações e Processamento de Sinais, 2020.

Abstract:

This paper presents an approach to target classification optimization based on acoustic signals collected using a hydrophone, an underwater electroacoustic transducer. This study has applications to sonars or any sound-classification application. We divide the problem into three parts, namely feature extraction, feature selection, and target classification with an optimization step. Experiments were conducted using ShipsEar, a public database of raw ship noises collected using a single hydrophone located in a harbor. This dataset comprises five classes and is used to verify the performance of the approach described in this work. From raw signals, we extracted the following features: Mel-Frequency Cepstral Coefficients, Linear Predictive Coding, and Gammatone Cepstral Coefficients. All these features were evaluated using the Neighborhood Component Analysis to reduce dimensionality. We used K-Nearest Neighbors as the classifier. We adopted the leave-one-out crossvalidation strategy to evaluate the classifier. Finally, we used Genetic Algorithms to optimize the features selected. We set the classifier performance as the genetic algorithm cost function and used the features selected as one individual of the first generation. This scheme optimized the performance of the classifier by 13 percentage points. In our case, the optimized feature selection algorithms reduced the dimensionality and improved classifier accuracy when compared with the same scheme using all features or a subset of features selected by Neighborhood Component Analysis. These techniques can select the most useful information from features of different ship classes.

Conference Paper:

2. FERNANDES, R. P.; APOLINÁRIO JR., J. A.; RAMOS, A. L. L.; SEIXAS, J. M. de. **Applying the majority voting rule in acoustic detection and classification of drones**. In: Sociedade Brasileira de Telecomunicações, 2021.

Abstract:

This paper discusses an approach to target detection and classification based on acoustic signals collected using one single microphone. We divide the problem into two parts, namely feature extraction and target detection and classification. We use an optimization step based on human auditory uncertainty. We employ a majority voting rule for every set of feature vectors, i.e., an estimate is only performed if the majority agrees. We conducted experiments using a single channel of the AIRA-UAS dataset, a public database of raw drone noises collected with an array of microphones mounted on a drone. This dataset comprises many different kinematics, with different spectra. The features we used are based on the Mel-Frequency Cepstral Coefficients (MFCC) and the Short-Time Fourier Transform of raw signals. We used the k-Nearest Neighbors algorithm for classification and adopted the cross-validation strategy to evaluate the method. We observed that the use of MFCC results in less biased estimations, which favors the voting strategy. The detection in the proposed method reached a probability of false positive near 0%, even with a small set of votes, and a classification accuracy of 99.1%. These metrics satisfy the requirements of most civilian and military applications.

Conference Paper:

3. FERNANDES, R. P.; APOLINÁRIO JR., J. A.; SEIXAS, J. M. de. **Enhancing TDE-based Drone DOA Estimation with Genetic Algorithms and Zero Cyclic Sum**. In: Congresso Brasileiro de Inteligência Computacional, 2023.

Abstract:

This paper discusses a way to enhance an acoustic-based approach to obtaining the direction of arrival (DOA) of a drone's ego noise using a microphone array. We focus on obtaining better time delay estimations (TDE) from a set of possible candidates. Recently, a large number of works have been put forward to detect and classify drones with different techniques. However, more investigation is required to tackle the drone DOA estimation problem using the time difference of arrival between pairs of microphones for the case of strongly corrupted audio signals, possibly by noise and multipath. The main problem in a complex acoustic environment is accurately estimating the time difference of arrival. With a traditional approach, this task becomes nearly impossible without the line of sight assumption, that is, whenever the highest cross-correlation peak between signals does not correspond to the delay between them. This paper uses genetic algorithms to search for the correct delays between pairs of microphones among a set of possible delays (primary and secondary delays). We define a fitness function based on the concept of zero cyclic sum of closed loops, i.e., when forming a closed loop, the sum of all theoretical delays should equal zero. A drawback of closed loops is that incorrect delays may result in a zero-sum; we thus created a fitness function that considers all possible closed loops of a given array. We exploited different approaches to estimate the direction of arrival using the combination of genetic algorithms and zero cyclic sum. In our experiments, the method successfully found all correct delays in simulations, providing strong evidence of its effectiveness when a correct delay exists among multiple possible delays. Furthermore, in experimental trials, it significantly enhanced the number of correct delays detected, further validating its utility and potential in practical scenarios.

Conference Paper:

4. APOLINÁRIO JR., J. A.; RAMOS, A. L. L.; FERNANDES, R. P.; DUARTE, J. C.; SOUSA, M. N. de. **Exploiting reverberation fingerprint for a neural network based acoustic emitter localization**. In: International Conference on Control, Decision and Information Technologies 2024.

Abstract:

This paper addresses the problem of localizing a sound source in strong reverberating and eventually noisy environments, including relevant potential applications, e.g., target location in interior room acoustic environments, in- and outdoor navigation, and robotics. Currently, available audio emitter location techniques, especially those based exclusively on Time Difference of Arrival, do not perform entirely well in unfriendly environments. Recent works propose using neural networks to solve the emitter location problem but have not yet adequately addressed the issue of severe reverberation. The solution proposed herein uses a two-layer feedforward neural network supported by signal processing techniques to extract the features employed in the neural network. Training the network with a select set of features, encompassing information from the reverberation fingerprint, results in efficient system featuring reduced training time and relatively low mean squared error values.

Journal Paper:

1. SERRENHO, F. G.; APOLINÁRIO JR., J. A.; RAMOS, A. L. L.; FERNANDES, R. P. **Gunshot airborne surveillance with rotary wing UAV-embedded microphone array**. *Journal on the Science and Technology of Sensors, Sensors*, v. 19(19), October 2019.

Abstract:

Unmanned aerial vehicles (UAV) are growing in popularity, and recent technological advances are fostering the development of new applications for these devices. This paper discusses the use of aerial drones as a platform for deploying a gunshot surveillance system based on an array of microphones. Notwithstanding the difficulties associated with the inherent additive noise from the rotating propellers, this application brings an important advantage: the possibility of estimating the shooter position solely based on the muzzle blast sound, with the support of a digital map of the terrain. This work focuses on direction-of-arrival (DOA) estimation methods applied to audio signals obtained from a microphone array aboard a flying drone. We investigate preprocessing and different DOA estimation techniques in order to obtain the setup that performs better for the application at hand. We use a combination of simulated and actual gunshot signals recorded using a microphone array mounted on a UAV. One of the key insights resulting from the field recordings is the importance of drone positioning, whereby all gunshots recorded in a region outside a cone open from the gun muzzle presented a hit rate close to 96%. Based on experimental results, we claim that reliable bearing estimates can be achieved using a microphone array mounted on a drone.

Journal Paper:

2. SOUSA, M. N. de; SANT'ANA, R.; FERNANDES, R. P.; DUARTE, J. C.; APOLINÁRIO JR., J. A.; THOMÄ, R. **Improving the performance of a radio-frequency localization system in adverse outdoor applications**. Journal on Wireless Communications and Networking, EURASIP, v. 1, p. 123, May 2021.

Abstract:

In outdoor RF localization systems, particularly where line of sight can not be guaranteed or where multipath effects are severe, information about the terrain may improve the position estimate's performance. Given the difficulties in obtaining real data, a ray-tracing fingerprint is a viable option. Nevertheless, although presenting good simulation results, the performance of systems trained with simulated features only suffer degradation when employed to process real-life data. This work intends to improve the localization accuracy when using ray-tracing fingerprints and a few field data obtained from an adverse environment where a large number of measurements is not an option. We employ a machine learning (ML) algorithm to explore the multipath information. We selected algorithms random forest and gradient boosting; both considered efficient tools in the literature. In a strict simulation scenario (simulated data for training, validating, and testing), we obtained the same good results found in the literature (error around 2 m). In a real-world system (simulated data for training, real data for validating and testing), both ML algorithms resulted in a mean positioning error around 100 ,m. We have also obtained experimental results for noisy (artificially added Gaussian noise) and mismatched (with a null subset of) features. From the simulations carried out in this work, our study revealed that enhancing the ML model with a few real-world data improves localization's overall performance. From the machine ML algorithms employed herein, we also observed that, under noisy conditions, the random forest algorithm achieved a slightly better result than the gradient boosting algorithm. However, they achieved similar results in a mismatch experiment. This work's practical implication is that multipath information, once rejected in old localization techniques, now represents a significant source of information whenever we have prior knowledge to train the ML algorithm.

Journal Paper:

3. FERNANDES, R. P.; APOLINÁRIO JR., J. A.; SEIXAS, J. M. de. **A reduced complexity acoustic-based 3D DOA estimation with Zero Cyclic Sum.** Journal on the Science and Technology of Sensors, Sensors, 2024.

Abstract:

Accurate Direction of Arrival (DOA) estimation is paramount in various fields, from surveillance and security to spatial audio processing. This work introduces an innovative approach that refines the DOA estimation process and demonstrates its applicability in diverse and critical domains. We propose a two-stage method that capitalizes on the often-overlooked secondary peaks of the cross-correlation function by introducing a reduced complexity DOA estimation method. In the first stage, we use a low complexity cost function based on zero cyclic sum (ZCS) condition that allows for an exhaustive search of all combinations of time delays between pairs of microphones, including primary peak and secondary peaks of each cross-correlation. For the second stage, we only test a subset of the time delay combinations with the lowest ZCS cost function using a least-squares (LS) solution, which requires more computational effort. To showcase the versatility and effectiveness of our method, we apply it to the challenging acoustic-based drone DOA estimation scenario using an array of four microphones. Through rigorous experimentation with simulated and actual data, our research underscores the potential of our proposed DOA estimation method as an alternative for handling complex acoustic scenarios. The ZCS method demonstrated an accuracy of $89.4\% \pm 2.7\%$, whereas the ZCS with LS method exhibited a notably higher accuracy of $94.0\% \pm 3.1\%$, showcasing the superior performance of the latter.

Journal Paper:

4. FERNANDES, R. P.; DUARTE, J. C.; APOLINÁRIO JR., J. A.; SEIXAS, J. M. de. **Optimized TDOA-based drone localization with distributed microphones.** Journal of Communication and Information Systems (**to be submitted**).

Abstract:

Accurately localizing drones in acoustically complex environments remains a significant challenge, with important implications for defense, law enforcement, and autonomous systems. This study addresses the problem of estimating drone localization using acoustics in environments characterized by strong reflections and noise. We employ Time Difference of Arrival (TDOA) techniques for localization estimation and compare them with a specialized machine learning regression model. While previous works have considered neural networks for localization, they often suffer from limited generalization across different environments. To address this, we propose a novel method that enhances the TDOA vector by incorporating both primary and secondary peaks of the cross-correlation, guided by the Zero Cyclic Sum condition. Additionally, we introduce optimization strategies that selectively reduce the number of TDOA inputs based on a least-squares cost function. We present a comparative analysis of TDOA-based optimization techniques with a machine learning method that utilizes the environment's reverberation fingerprint as input features for training. Experimental results demonstrate that the proposed TDOA-based method achieves a localization accuracy of 0.55 ± 0.35 meters, showcasing its effectiveness and practical applicability in challenging acoustic environments.